THE PROCEEDINGS OF
THE PHYSICAL SOCIETY
Section B

VOL. 66, PART 1 1 January 1953 No. 397 B

EDITORIAL

The 1952 Volumes of the *Proceedings of the Physical Society* were about 7% smaller in size than those of 1951, although the number of papers published increased by some 5%. In the Editorial of January 1952 the need for greater conciseness was stressed, and this, together with the patience and vigilance of referees, may account for the reduction in the average length of papers.

The Papers Committee considered during the year an analysis of the papers published in 1950 and 1951. It was reported that the average time required to publish papers in 1950 was 6½ months. This had decreased to 5½ months in 1951, and it is pleasing to record that a preliminary analysis for 1952 shows a further decrease to 5 months. Nevertheless the Committee feels that this is too long—even for the average—and steps are in hand which, it is hoped, will reduce this period. Authors can help greatly in this matter by paying attention to the preparation of manuscripts, and by submitting duplicate copies of the text together with three copies of the abstract and rough copies of diagrams.

A small sub-committee considered, and reported on, the quality of Letters to the Editor. The position was considered to be reasonably satisfactory and no action called for.

Beginning in January 1953, it is proposed to accept Research Notes for publication. These should be shorter in form than a Paper, and are intended for the communication of results where the work reported does not need a full paper.

During the latter part of the year Dr. E. P. George has resumed his work for the Society in assisting me in the consideration of papers submitted for publication in the *Proceedings*. Professor S. Devons continued to act as Chairman of the Papers Committee.

H. H. HOPKINS,
Honorary Papers Secretary.

The Mean Shear Stress in an Array of Dislocations and Latent Hardening

BY A. N. STROH

H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott; MS. received 28th July 1952, *and in final form* 14*th October* 1952

ABSTRACT. The root mean square shear stress on a slip plane due to an array of dislocations is taken as a measure of the work hardening of that plane. The ratio of the hardening of the latent slip system to that of the operative system, which is shown to be independent of the distribution of the dislocations, is calculated both for screw and for edge dislocations. Numerical values are given for the twelve slip systems of face-centred cubic crystals. The results suggest that hardening by edges may account for the greater part of the observed latent hardening.

§ 1. INTRODUCTION

In the model of work hardening of metals which was proposed by Taylor (1934), hardening is attributed to the effects of the stress field of an array of dislocations lying in the operative slip planes. This is also a feature of Mott's theory of work hardening (1952) which differs from the original Taylor theory in that the dislocations are now supposed to pile up at obstacles, instead of being distributed regularly throughout the crystal. The purpose of the present paper is to investigate what hardening these dislocations will produce on the latent slip systems; in terms of the model this is equivalent to asking how the mean shear stress will vary with direction. We shall not find it necessary to specify the distribution of the dislocations, so that our results will still apply if a pile up of dislocations occurs.

However, it should be noted that other mechanisms for the production of work hardening have also been proposed; for instance, Cottrell (1951) has suggested that the essential process is the work done in forcing two mutually perpendicular dislocations past one another. Such theories may be expected to predict quite different latent hardening and will not be considered here. A different treatment of the problem has been given recently by Röhm and Diehl (1952).

We choose coordinates (x, y, z) such that the xy plane is the operative slip plane and the x axis the operative slip direction; and coordinates (ξ, η, ζ), with the $\xi\eta$ plane the latent slip plane and the ξ axis the latent slip direction. $l_{x\xi}, l_{x\eta}, \ldots$ are the direction cosines of the ξ, η, ζ axes referred to the x, y, z axes.

§ 2. HARDENING DUE TO SCREW DISLOCATIONS

First we shall consider the case in which all the dislocations are screws; in our coordinates these will be parallel to the x axis. The stresses at the origin due to a single dislocation through the point $(0, y, z)$ are

$$\tau_{xy} = -\frac{Gb}{2\pi}\frac{\sin\theta}{r}, \qquad \tau_{xz} = \frac{Gb}{2\pi}\frac{\cos\theta}{r}, \qquad \ldots\ldots(1)$$

where $r^2 = y^2 + z^2$ and $\tan\theta = z/y$, all other stress components vanishing. G is the rigidity modulus and b Burgers' vector.

Now

$$\tau_{\xi\eta} = (l_{x\xi}l_{y\eta} + l_{x\eta}l_{y\xi})\tau_{xy} + (l_{x\xi}l_{z\eta} + l_{x\eta}l_{z\xi})\tau_{xz}, \qquad \ldots\ldots(2)$$

so that on substituting from (1) in (2) we have

$$\tau_{\xi\eta} = \frac{Gb}{2\pi}\,\frac{-m_1\sin\theta + m_2\cos\theta}{r}, \qquad \ldots\ldots(3)$$

where

$$m_1 = l_{x\xi}l_{y\eta} + l_{x\eta}l_{y\xi}, \qquad m_2 = l_{x\xi}l_{z\eta} + l_{x\eta}l_{z\xi}. \qquad \ldots\ldots(4)$$

Suppose that the position of the ith dislocation is given by (r_i, θ_i). Then the resultant stress of the whole array is

$$\tau_{\xi\eta} = \sum_i \pm \frac{Gb}{2\pi}\frac{1}{r_i}(-m_1\sin\theta_i + m_2\cos\theta_i), \qquad \ldots\ldots(5)$$

where the positive or negative sign must be taken according as the ith dislocation is positive or negative.

Changing the point of observation (the origin of the coordinates) will alter the value of θ_i, and since every term in (5) contributes numerically equal positive and negative values to the sum as θ_i varies from 0 to 2π, the stress will fluctuate about the mean value zero. Accordingly we consider the mean square shear stress.

$$\begin{aligned}\tau_{\xi\eta}{}^2 = {}& \frac{G^2b^2}{4\pi^2}\sum_i \frac{1}{r_i{}^2}(m_1{}^2\sin^2\theta_i - m_1m_2\sin 2\theta_i + m_2{}^2\cos^2\theta_i)\\ &+ \frac{G^2b^2}{4\pi^2}\sum_{i>j} \pm \frac{1}{r_ir_j}\{(m_1{}^2+m_2{}^2)\cos(\theta_i-\theta_j)\\ &-(m_1{}^2+m_2{}^2)\cos(\theta_i+\theta_j) - 2m_1m_2\sin(\theta_i+\theta_j)\}. \qquad \ldots\ldots(6)\end{aligned}$$

Now clearly all values of θ_i, and of $\frac{1}{2}(\theta_i+\theta_j)$ the mean coordinate of a pair of dislocations, are equally probable; so that on averaging eqn. (6) we may replace $\sin^2\theta_i$ and $\cos^2\theta_i$ by $\frac{1}{2}$, while the terms involving $\sin 2\theta_i$, $\sin(\theta_i+\theta_j)$ and $\cos(\theta_i+\theta_j)$ vanish. If the dislocations are distributed at random then it is also true that all values of θ_{ij} (defined by $\theta_{ij} = \theta_i - \theta_j$), are equally probable; but if the dislocations are piled up in groups, which seems likely to be the case (Mott 1952), then small values of θ_{ij} will occur more frequently and we cannot find the mean explicitly without specifying the distribution in greater detail. Denoting the mean of $\tau_{\xi\eta}{}^2$ by $\bar{\tau}_{\xi\eta}{}^2$ and also using a bar for other mean values, we have

$$\bar{\tau}_{\xi\eta}{}^2 = \frac{G^2b^2}{8\pi^2}\left\{\overline{\sum_i \frac{1}{r_i{}^2} + 2\sum_{i>} \pm \frac{\cos\theta_{ij}}{r_ir_j}}\right\}(m_1{}^2+m_2{}^2), \qquad \ldots\ldots(7)$$

or using (4)

$$(\bar{\tau}_{\xi\eta}/\tau_0)^2 = l_{x\xi}{}^2 + l_{x\eta}{}^2 - 4l_{x\xi}{}^2l_{x\eta}{}^2, \qquad \ldots\ldots(8)$$

where

$$\tau_0{}^2 = \frac{G^2b^2}{8\pi^2}\left\{\overline{\sum_i \frac{1}{r_i{}^2} + 2\sum_{i>j} \pm \frac{\cos\theta_{ij}}{r_ir_j}}\right\}. \qquad \ldots\ldots(9)$$

τ_0 is the mean shear stress on the operative slip system. Equation (8) shows that the ratio of the mean stress on the two slip systems is independent of the distribution of the dislocations and depends only on the relative orientation of the two systems.

§ 3. EDGE DISLOCATIONS

If the hardening is supposed due to edge dislocations, these must be taken parallel to the z axis and with Burgers' vector along the x-axis. (Burgers' vector gives the slip associated with each dislocation.) The stresses due to a single dislocation are

$$\left.\begin{aligned}
\tau_{xx} &= \frac{Dy(3x^2+y^2)}{(x^2+y^2)^2} = -\frac{D\sin\theta\cos 2\theta}{r} - \frac{2D\sin\theta}{r}, \\
\tau_{yy} &= \frac{Dy(x^2-y^2)}{(x^2+y^2)^2} = \frac{D\sin\theta\cos 2\theta}{r}, \\
\tau_{zz} &= \sigma(\tau_{xx}+\tau_{yy}) = -\frac{2\sigma D\sin\theta}{r}, \\
\tau_{xy} &= \frac{Dx(x^2-y^2)}{(x^2+y^2)^2} = \frac{D\cos\theta\cos 2\theta}{r}, \\
\tau_{yz} &= \tau_{xz} = 0,
\end{aligned}\right\} \qquad \ldots\ldots(10)$$

where $D = bG/2\pi(1-\sigma)$, σ is Poisson's ratio, $r^2 = x^2+y^2$, $\tan\theta = y/x$. These stresses give

$$\tau_{\xi\eta} = D\Sigma \pm (n_1 \sin\theta_i \cos 2\theta_i - 2n_2 \sin\theta_i + n_3 \cos\theta_i \cos 2\theta_i)/r_i, \qquad \ldots\ldots(11)$$

where

$$n_1 = l_{y\xi} l_{y\eta} - l_{x\xi} l_{x\eta}, \quad n_2 = l_{x\xi} l_{x\eta} + \sigma l_{z\xi} l_{z\eta}, \quad n_3 = l_{x\xi} l_{y\eta} + l_{x\eta} l_{y\xi}. \qquad \ldots\ldots(12)$$

On squaring (11) and taking the mean as before, we have

$$\bar{\tau}_{\xi\eta}{}^2 = (n_1{}^2 + n_3{}^2)\tau_0{}^2 + 4(2n_2{}^2 + n_1 n_2)\tau_1{}^2 \qquad \ldots\ldots(13)$$

where

$$\tau_0{}^2 = \frac{D^2}{4}\left(\overline{\sum_i \frac{1}{r_i{}^2}} + 2\overline{\sum_{i>j} \pm \frac{\cos\theta_{ij}\cos 2\theta_{ij}}{r_i r_j}}\right), \qquad \ldots\ldots(14)$$

and

$$\tau_1{}^2 = \frac{D^2}{4}\left(\overline{\sum_i \frac{1}{r_i{}^2}} + 2\overline{\sum_{i>j} \pm \frac{\cos\theta_{ij}}{r_i r_j}}\right). \qquad \ldots\ldots(15)$$

τ_0 is the shear stress on the operative slip system. In the case in which the dislocations are distributed at random the second summations in (14) and (15) are zero and $\tau_1 = \tau_0$. It seems likely that $\tau_1 = \tau_0$ will still be a good approximation when the pile up of dislocations occurs. For the terms occurring in the second summations in (14) and (15) may be divided into two sets according as the ith and jth dislocation are in the same group or in different groups of piled up dislocations. If they come from different groups we regard them as having a random distribution and consequently making a net contribution of zero to the summation, while if they are in the same group, θ_{ij} will be small in all but a few cases and we may replace the cosines by unity without too great an error. The expressions (14) and (15) then give $\tau_1 = \tau_0$. Further, the effect of any error made by putting $\tau_1 = \tau_0$ in (13) will be reduced by the fact that the coefficient of $\tau_1{}^2$ is generally found to be somewhat smaller than that of $\tau_0{}^2$. Making this approximation, we have from (13) and (12):

$$(\bar{\tau}_{\xi\eta}/\tau_0)^2 = (1 - l_{z\xi}{}^2)(1 - l_{z\eta}{}^2) - 4(1-2\sigma) l_{z\xi} l_{z\eta}. \qquad \ldots\ldots(16)$$

Equation (16) gives the mean stress due to edge dislocation and corresponds to eqn. (8) for screws.

§4. DISCUSSION OF RESULTS

From the crystallographic indices of the slip planes and directions, it is a simple matter to calculate the direction cosines occurring in (8) and (16), and so to find $\bar{\tau}_{\xi\eta}/\tau_0$. For edges the values obtained will depend on the value of Poisson's ratio assumed, though this dependence is in most cases small. Results for the twelve slip systems of the face-centred cubic lattice are given in the table. Here the operative slip system has been taken as the $[01\bar{1}]$ direction in the (111) plane.

Slip plane	Slip direction	Values of $100\,\bar{\tau}_{\xi\eta}/\tau_0$			
		Screws	Edges		
			$\sigma=0$	$\sigma=\frac{1}{4}$	$\sigma=\frac{1}{2}$
(111)	$[01\bar{1}]$	100	100	100	100
	$[\bar{1}01]$	50	50	50	50
	$[1\bar{1}0]$	50	50	50	50
$(\bar{1}11)$	$[01\bar{1}]$	100	33·3	33·3	33·3
	[101]	50	31·9	25·2	31·9
	[110]	50	31·9	25·2	31·9
$(1\bar{1}1)$	[011]	81·7	72·0	69·3	72·0
	$[\bar{1}01]$	50	92·8	66·7	44·1
	[110]	50	96·8	90·2	84·5
$(11\bar{1})$	[011]	81·7	72·0	69·3	72·0
	[101]	50	96·8	90·2	84·5
	$[1\bar{1}0]$	50	92·8	67·7	44·1

The experiments of Taylor and Elam (1925), Elam (1926) and Göler and Sachs (1929) show that when a second slip system is brought into operation through rotation of the lattice, this system is as hard as, or slightly harder than, the operative system. If the operative system is, as above, the $[01\bar{1}]$ direction in the (111) plane, then the new slip system will be the [110] direction in the $(1\bar{1}1)$ plane or the [101] direction in the $(11\bar{1})$ plane. From the table we see that the predicted hardening is 50% for screws (which is much too low), and about 90% for edges. Thus hardening by edges alone could account for the greater part, but not the whole, of the observed effect. It is to edges that the hardening is attributed both in the theory of Taylor (1934) and in that of Mott (1952).

In order to explain the observed fine structure of the slip bands in aluminium which has been resolved by the electron microscope, Brown (1951) and Mott (1952) have supposed that some self-annealing takes place in the operative slip planes. This would result in increasing the apparent hardness of the latent slip system relative to that of the operative, and consequently might be responsible for the difference between our predicted hardening and that observed experimentally. The experiments were carried out at room temperature, and since the self-annealing is presumably temperature dependent, we should expect, if this interpretation is correct, that at sufficiently low temperatures the latent hardening would decrease to our theoretical value. That is, if a metal is deformed plastically at liquid air temperatures, and an appreciable amount of slip has taken place on a single system, then the second system should become operative slightly before the lattice has been rotated to the symmetrical position. This would be the best test of the theory presented here. However, since existing work has been done mainly on aluminium, useful results might also be obtained even at room temperature from metals such as nickel which have different activation energies.

From the table we see that in some cases the latent slip system may be only about a third as hard as the operative. After a reasonable amount of deformation has taken place the difference in hardness should be quite appreciable. Röhm and Kochendörfer (1950) have given results showing that the latent slip systems are hardened less than the operative and that the amount of hardening depends on the particular slip system considered. An indirect confirmation is provided by the phenomenon of cross slip which is observed to accompany slip on the operative system. The theory presented here leads to a simple interpretation of this, for whether or not slip takes place on any slip plane depends both on the resolved shear stress and on the amount of work hardening of that plane. The resolved shear stress on a secondary system is less than that on the operative, but as the work hardening is also less it is still possible that the stress on the secondary system may be great enough to produce slip. This would be observed as cross slip. The presence of some cross slip would increase the rate at which the secondary system is hardened and this would then tend to prevent further cross slip, thus limiting its amount to a fraction of that on the primary system.

REFERENCES

Brown, A. F., 1951, *J. Inst. Metals*, **19**, 115.
Cottrell, H., 1951, Report of the Solvay Conference, p. 421.
Elam, C. F., 1926, *Proc. Roy. Soc.* A, **112**, 289.
Göler, v., and Sachs, G., 1929, *Z. Phys.*, **55**, 581.
Mott, N. F., 1952, *Phil. Mag.*, **43**, 1151.
Röhm, F., and Diehl, J., 1952, *Z. Metallkunde*, **43**, 126.
Röhm, F., and Kochendörfer, A., 1950, *Z. Metallkunde*, **41**, 265.
Taylor, G. I., 1934, *Proc. Roy. Soc.* A, **145**, 362.
Taylor, G. I., and Elam, C. F., 1925, *Proc. Roy. Soc.* A, **108**, 282.

Crystal Structures of Gutta Percha

By DOROTHY FISHER

British Rayon Research Association, Manchester

MS. received 22nd August 1952

ABSTRACT. Thin films of gutta percha have been examined by means of electron diffraction, and shown to exist in three distinct crystalline forms, termed the α, β and γ modifications, whereas prior to this work only two forms had been recognized.

The conditions necessary for the occurrrence of the different forms are considered and probable unit cells are suggested for two of them. These have the dimensions :

β form : $a=7{\cdot}83$ Å, $b=11{\cdot}87$ Å, $c=4{\cdot}75$ Å, $\alpha=\beta=\gamma=90°$.
γ form : $a=5{\cdot}9$ Å, $b=7{\cdot}9$ Å, $c=9{\cdot}2$ Å, $\alpha=\beta=90°$, $\gamma=94°$.

Data from the α modification were insufficient to provide the full set of lattice constants but the identity period along the chain axis was found to be 8·76 Å.

The β form which provides the most data has received much previous study and the present figures show broad agreement with those of previous workers.

The three identity periods of 8·76 Å, 4·75 Å, and 9·2 Å are in close agreement with those predicted by C. W. Bunn on the basis of theoretical considerations of the possible configurations of a transpolyisoprene chain.

§ 1. INTRODUCTION

GUTTA PERCHA is a naturally occurring polymer of isoprene characterized by the fact that the isoprene units are in the *trans* form in relation to the double bonds. In this respect it differs from natural rubber in which the isoprene units are in the *cis* configuration (Meyer and Mark 1930). It shows considerable crystallinity in the unstretched state at room temperature, and earlier work by X-ray and electron diffraction has established the existence of two crystal forms which were termed the α and β modifications (Hopff and von Susich 1930, Hauser and von Susich 1931). The β structure predominates in most commercial forms of gutta percha and has consequently received the most study. Bunn (1942) has proposed a detailed atomic structure based on an orthorhombic cell with $a=7{\cdot}85$ Å, $b=11{\cdot}9$ Å, $c=4{\cdot}72$ Å.

In a discussion of the possible configurations of a *trans* polyisoprene chain, Bunn (1942) predicted three possible structures for gutta percha which he designated α, β, γ. It appears that previous experimental work has failed to distinguish between the α and γ structures and consequently the measurements of different workers purporting to give the lattice spacings of the α form do not agree with one another. In the present work the existence of three forms is clearly shown and the conditions for their occurrence are partially defined.

§ 2. EXPERIMENTAL

The material used was a commercially purified grade of gutta percha. Specimens were prepared in the form of thin films suitable for transmission work by depositing them from benzene solution on to a clean water surface in a Langmuir trough. The solution spreads readily on the water surface thus providing an extremely thin film of gutta percha after evaporation of the solvent.

It was necessary to use freshly made solutions and to examine the films immediately after their preparation since the material oxidizes very rapidly.

Some films were examined in the unstretched condition; others were stretched by placing two chromium bars across the film in the trough and then drawing them apart. The films were somewhat brittle and inclined to tear under this treatment but the extension was helped by warming the substrate water to about 30° c and it was then found possible to stretch them to about four times their original length. Films examined in the unstretched state had a thickness of about 200–300 Å as indicated by the diffraction colours under oblique illumination, while those used for stretched specimens had a thickness of about 1000 Å before stretching. Portions of the film were gathered on clean specimen grids and then examined in a Finch-type electron diffraction camera working with a camera length of 47 cm and an accelerating voltage of 50–60 kv.

§ 3. EXPERIMENTAL RESULTS

Although selected from the same grade of bulk material the specimens examined were found to be of two distinct kinds whose diffraction patterns are illustrated in figs. 1 and 2 (Plate).

One group (fig. 1) existed in the well-known β modification and gave its characteristic diffraction pattern. The principal effect of stretching these samples was a change in the orientation of the crystallites, although traces of a subsidiary component, presumably the α-form, could be observed in the patterns from stretched samples.

The second group (fig. 2) behaved in a more complex manner in that stretching these samples produced a complete change of structure. Stretched samples exhibited the same characteristics as those in the first group, being composed predominantly of β material with traces of α structure, but unstretched samples gave rise to a pattern which differed from either the α or β pattern, and appeared to arise from a third crystal form which we have termed the γ form. Since only two crystal forms for gutta percha had been reported it was at first thought possible that the α and γ patterns corresponded to differently orientated arrangements of the same crystalline structure, and attempts were made to index the spacings accordingly, but it was soon found that this was not possible and that two distinct structures were involved in addition to the β structure. It is in relation to these two α and γ structures that the principal interest of the present work lies.

A study of the literature leaves little doubt that both these structures have been observed previously; thus the identity period of 8·76 Å observed in the present work for the α material corresponds to that reported for α gutta percha by Hauser and von Susich (1931) who worked with stretched specimens giving well oriented fibre diagrams, and later observed by Fuller (1936), while on the other hand the principal spacings in the ring pattern are in close agreement with those previously reported for unstretched specimens by Hopff and von Susich (1930) and Hauser and von Susich (1931), by Stilwell and Clark (1931) and by Bruni and Natta (1934) who worked with electron diffraction. But owing to the fact that the x-ray patterns have been relatively poorly developed and frequently observed in association with the β pattern, workers in this field do not appear to have attempted to derive the complete set of lattice constants. Consequently the discrepancies between stretched and unstretched samples have not been appreciated and both structures have been loosley termed α gutta percha, although it is true that Bruni and Natta working with unstretched films defined a unit cell with an identity period different from that observed in the oriented X-ray patterns, and that Fuller observed certain discrepancies in the identity period as determined from different so-called α reflections which led him to suspect the possible existence of a third form of gutta percha.

The terminology adopted here has been selected so as to conflict as little as possible with that previously used. Some conflict is obviously unavoidable but the majority of workers appear to have accepted the identity period of 8·7–8·8 Å as characteristic of the α modification. The structure appearing as a subsidiary component in stretched samples in the present work exhibits this identity period and has therefore been termed the α form. The structure occurring in unstretched samples (other than β specimens) then becomes the γ modification, notwithstanding the fact that present evidence suggests that it is the form in which the material exists when first drawn from the tree. This nomenclature is in conformity with that used by Bunn.

The conditions governing the formation of the different crystalline modifications were not fully investigated in the present experiments but certain major conditions for their occurrence in thin films suitable for electron diffraction were observed. The bulk material from which the test specimens were selected was a commercially purified grade of gutta percha which had presumably suffered heat treatment in the course of its processing. When first examined all unstretched specimens exhibited the β structure while, on stretching, the β structure remained predominant but a small admixture of α material became apparent. When further samples were selected some months later many unstretched specimens exhibited the γ structure but stretched specimens still showed the same characteristics as before. It was found however, that if the solid material were maintained at 80–90° C for about half an hour immediately before forming the solution from which the test specimen was deposited, the appearance of the γ structure was avoided and all unstretched specimens exhibited the β structure.

Summarizing these observations it may be said that at room temperature unstretched specimens exist in either the β or the γ modification according to the previous heat treatment of the material, but that on stretching, all samples revert to a mixture of the α and β modifications in which the β component predominates.

These conclusions refer only to thin films of unvulcanized material; the conditions for their appearance in bulk samples of vulcanized material may be very different, indeed Fuller (1936) who was largely concerned with a study of the effects of vulcanization quotes spacings agreeing closely with the present γ spacings, as the equatorial spacings in fibre diagrams of stretched material thus suggesting that under the condition of his experiments the γ structure persists on stretching. However, in so far as they relate to unstretched specimens these observations are in broad agreement with the indications in the literature that the γ modification, termed α in earlier work, is genuinely stable at room temperature, but that the β form, developed at temperatures above about 70° C, is retained on normal cooling and exists in a metastable state at room temperature owing to its low transformation rate back to the γ form at temperatures below the transition point. The fact that a proportion of the material appeared to have reverted to the γ form during storage cannot be fully explained since the conditions of commercial processing were not known in detail. It is probable, however, that the transition was assisted by the process of solution and recrystallization involved in forming the test specimens for the present work since a change of structure under this treatment was observed by Hopff and von Susich (1930). The reversion to a material consisting largely of the β modification on stretching γ samples has also been observed previously (Hauser 1927) but at that time the polymorphism of gutta percha was not understood.

The recognition of the subsidiary α component in stretched samples as having a distinct structure different from the γ structure in unstretched specimens does not appear to have been reported before, and this, together with the further definition of these two structures, provides the principal contribution of this work to our knowledge of gutta percha.

§ 4. DETAILS OF SPACINGS AND PROPOSED UNIT CELLS

Details of the spacings observed are given in tables 1, 2 and 3 together with comparable published data where this is available. As remarked previously,

Table 1. β Gutta Percha (figs. 1 (*a*), (*b*) and (*c*) and 2 (*c*))

Proposed cell: Orthorhombic, $a=7\cdot84$ Å, $b=11\cdot87$ Å, $c=4\cdot75$ Å

(1)	(2)	(3)	(4)			
			(i)	(ii)	(iii)	(iv)
4·73 vs	120	4·73	4·75 vs, 4·73 m	4·8	4·73 vvs	4·71
3·92 vs	200	3·92	3·90 m, 3·89 s	3·9	3·91 vs	3·87
2·98 mw	040	2·97	2·95 m		3·15 vw	3·32
2·79 mw	140, 230	2·79, 2·78	2·77 w		2·98 m	2·96
2·38 m	320	2·39			2·78 s	2·75
1·97 mw	400, 340	1·96, 1·96			2·37 s	2·38
1·93 m	410	1·93			2·12 vw	2·17
1·77 w	430	1·75			2·02 vw	2·04
4·44	011	4·41			1·95 m	1·97
2·98	211, 201	2·93, 3·03				1·91
2·26	311, 301	2·24, 2·28				1·76
1·80	411, 401, 341	1·79, 1·81, 1·81				1·18
2·38	002	2·37				
1·57	003	1·58				
1·52	113	1·54				
1·19	004	1·18				

(1) Observed spacing (Å) relative to graphite (110) = 1·230 Å; (2) Indices suggested (*hkl*); (3) Spacings (Å) computed for $a=7\cdot84$ Å, $b=11\cdot87$ Å, $c=4\cdot75$ Å; (4) Published figures Å: (i) Hopff and von Susich (1930), (ii) Hauser and von Susich (1931), (iii) Fuller (1936), (iv) Storks (1938) (electron diffraction).

s = strong, m = medium, w = weak, vs = very strong, etc.

Note.—This table gives complete set of spacings from stretched and unstretched specimens. Unstretched specimens when viewed perpendicular to the beam showed only spacings of (*hk*0) type.

Table 2. α Gutta Percha
(Subsidiary pattern in figs. 1 (*c*) and 2 (*c*)—stretched samples)

Proposed cell : data are insufficient for determination of full set of lattice constants, but layer line spacing gives a value of 8·76 Å for the *c* axis

Layer line	0	1	2	4	5	6
Observed spacing (Å)*	3·36	4·55 4·07 3·36 2·73	3·49	2·18 2·10	1·57	1·46

Published figures : side spacings are not in general quoted in detail but Hauser and von Susich agree with Fuller on an identity period of about 8·8 Å corresponding to the value of 8·76 Å for the *c* axis in the present work.

* Relative to graphite (110) = 1·230 Å.

the term α gutta percha has been used in the literature for any modification differing from the β form, but for these tables the published figures have been reclassified as relating to the α or γ forms according to the condition in which the sample was examined and the results obtained.

Table 3. γ Gutta Percha (figs. 2 (*a*) and (*b*)—unstretched samples)

Proposed cell : Monoclinic $a=5{\cdot}9$ Å, $b=7{\cdot}9$ Å, $c=9{\cdot}2$ Å, $\alpha=\beta=90°$, $\gamma=94°$

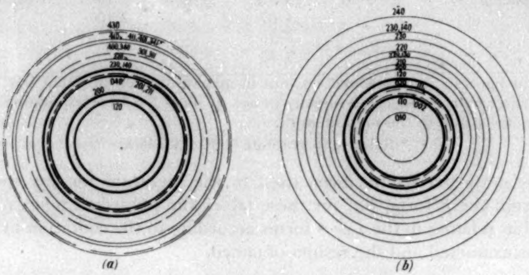
(1)	(2)	(3)	(4)		
			(i)	(ii)	(iii)
7·91 m	010	7·90		7·7	12·2
4·95 vs	$1\bar{1}0$	4·88	4·82		4·97, 4·96, 4·56
3·95 vs	020	3·95	3·86	3·9	3·94
3·36 vs	$1\bar{2}0$	3·39	3·28	3·3	3·32
2·95 mw	200	2·95	2·95		2·98
2·73 mw	210	2·70	2·69		2·74
2·47 vw	$2\bar{2}0$, $1\bar{3}0$	2·44, 2·46	2·48, 2·42		
2·29 w	220	2·28	2·22		
2·03 vw	$2\bar{3}0$	2·03	2·04		
1·89 ms	230, $1\bar{4}0$	1·89, 1·91	1·90		
1·65 m	240, 240	1·69, 1·59	1·64		
			1·24		
4·56*	002	4·60			
4·08*	111	4·09			

(1) Observed spacing (Å) relative to graphite (110) = 1·230 Å; (2) Suggested indices (*hkl*); (3) Spacing (Å) computed for $a=5{\cdot}9$ Å, $b=7{\cdot}9$ Å, $c=9{\cdot}2$ Å, $\gamma=94°$; (4) Published results (Å) for modification (referred to in earlier literature as unstretched α material) : (i) Bruni and Natta 1934, (ii) Hauser and von Susich 1931, (iii) Hopff and von Susich 1930.

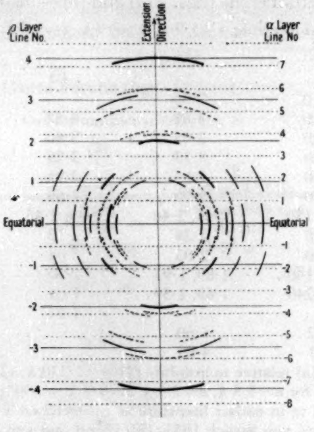
* Specimen perpendicular to beam except in these two cases. Here specimen is inclined at 45° to beam. Spacings along the diameter parallel to the axis of inclination were unchanged. Two additional arcs appeared on the perpendicular diameter at positions stated.

Absolute values of the spacings were determined from specimens (not illustrated) in which a thin layer of colloidal graphite was deposited on the gutta percha so that a composite pattern was obtained; the spacings were then computed relative to the standard graphite (110) spacing of 1·230 Å (Finch and Fordham 1936, Trzebiatowski 1937, Nelson and Riley 1945). Figures for the β material were included for the sake of completeness and to indicate the order of agreement between this and previous work.

Diagrammatic representations of the three principal types of pattern are given in figs. 3 (*a*), (*b*), (*c*), and where possible indicate the indices assigned to the various reflections. Figure 3 (*c*), relating to the composite pattern for stretched samples, also shows how the observed arcs are distributed between the two component structures.


430
230,140
040
200
120
240
230,140
230
220
210
120
110
002
040

(*a*) (*b*)


β Layer Line No.
Extension Direction
α Layer Line No
4
3
2
1
Equatorial
-1
-2
-3
-4
7
6
5
4
3
2
1
Equatorial
-1
-2
-3
-4
-5
-6
-7
-8

(*c*)

Fig. 3 (*a*) and (*b*). Patterns from unstretched gutta percha (diagrammatic). Broken circles indicate arcs which appear on inclining film to beam. (*a*) β gutta percha; (*b*) γ gutta percha.

Fig. 3 (*c*). Combined α and β patterns (diagrammatic) from stretched gutta percha. α arcs shown dotted.

β *Structure*

The data from the β modification is more comprehensive than that from the other two types since it exists in both the stretched and unstretched state and therefore provides patterns showing crystallites in two different orientations. In interpreting these patterns the lattice constants proposed by Bunn (1942) and by Fuller (1936) were used as guides for indexing the spacings and the axial lengths were then adjusted to agree with the observed radii. The unit cell thus determined was orthorhombic with $a = 7{\cdot}83$ Å, $b = 11{\cdot}87$ Å, c (fibre axis) $= 4{\cdot}75$ Å, which, taking into account the fact that some of the arcs were rather broad, seems to be in satisfactory agreement with Bunn's cell, having $a=7{\cdot}85$ Å, $b=11{\cdot}9$ Å, and $c=4{\cdot}72$ Å.

In the unstretched material there is pronounced orientation of the crystallites with their chain axes in a direction perpendicular to the plane of the film, while on stretching, the chain axis tends to take the direction of extension giving rise to the familiar fibre or partial rotation diagram. The former is clearly indicated by the pronounced arcing which occurs on inclining the film to the electron beam (fig. 1 (*b*)) and by the fact that the reflections occurring when the film is perpendicular to the beam are confined to those of the (*hk*0) type. The same phenomenon was observed by Storks (1938) and is, as he pointed out, somewhat surprising since the thickness of the film (~300 Å) is several times smaller than

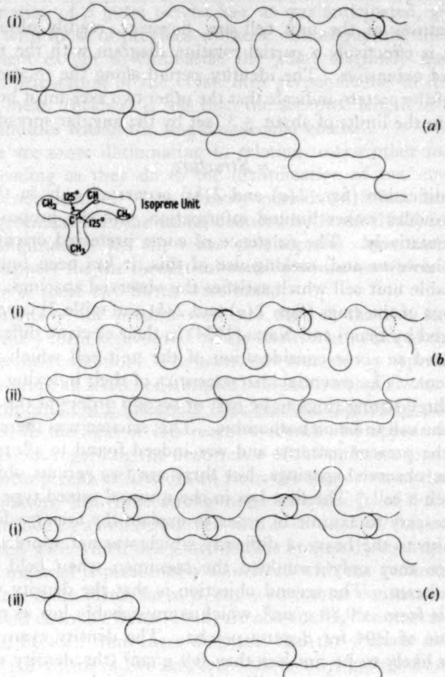


Fig. 4. Gutta percha molecules (diagrammatic) according to C. W. Bunn. (*a*) α, (*b*) β, (*c*) γ.
(i) Projection in plane of double bonds.
(ii) Projection perpendicular to plane of double bonds.

the length of a single molecular chain (~3000 Å) as determined by independent means and thus appears to indicate sharp discontinuities in the chain.

α *Structure*

The number of arcs available for α gutta percha patterns (figs. 1 (*c*) and 2 (*c*)) is relatively few, and the fact that they are superimposed on the β pattern which exhibits several similar spacings introduces some ambiguity. Consequently it has not been found possible at this stage to deduce a satisfactory unit cell for this material.

Certain features of the unit cell are, however, readily obtained from the pattern which is effectively a partial rotation diagram with the rotation axis in the direction of extension. The identity period along the rotation axis (8·76 Å) and the form of the pattern indicate that the other two axes must be perpendicular to this to within the limits of about ± 3° set by the angular spread of the arcs.

γ *Structure*

The γ modification (figs. 2 (*a*) and 2 (*b*)) occurring only in the unstretched state, again provides rather limited information, and the proposed unit cell is put forward tentatively. The existence of some preferred orientation aids the interpretation however and, making use of this, it has been found possible to deduce a probable unit cell which satisfies the observed spacings.

The spacings of the rings (figs. 2 (*a*) and 2 (*b*) and table 3) correspond closely to those observed by Bruni and Natta (1934) in their electron diffraction study of gutta percha, and so some consideration of the unit cell which they proposed seemed expedient. The essential characteristics of their indexing were that they supposed the three strong rings to be first or second orders of the principal axial spacings, and the cell to be orthorhombic. This scheme was therefore examined in relation to the present patterns and was indeed found to afford a neat interpretation of the observed spacings, but there are two serious objections to the adoption of such a cell. The first lies in the unusual mixed type of orientation which it is necessary to assume in order to explain the arcing which occurs on inclining the film to the beam, a difficulty which was not appreciated by Bruni and Natta since they only examined the specimen when held perpendicular to the electron beam. The second objection is that the density value deduced for a cell of this form is 0·86 g/cm^3, which is improbably low as compared with the known value of 1·04 for β gutta percha. The density even of amorphous gutta percha is likely to be not less than 0·9 g/cm^3 (the density of unstretched amorphous rubber), and all crystalline forms would be expected to be more dense than the amorphous material.

An alternative cell was therefore sought, and the observed spacings were found to fit a monoclinic cell with $a = 5{\cdot}9$ Å, $b = 7{\cdot}9$ Å, c (fibre axis) $= 9{\cdot}2$ Å, $\alpha = \beta = 90°$, $\gamma = 94°$.

On the basis of this cell, all the rings appearing when the film is examined perpendicular to the electron beam are indexed as ($hk0$) spacings and the orientation is thus seen to be similar to that occurring in the unstretched β material, and is such that the fibre axis lies perpendicular to the plane of the film.

The length of the *c* axis is derived only from the two additional arcs which appear when the specimen is inclined to the beam, and is thus subject to some uncertainty. The figure quoted is derived on the assumption that the *c* axis is perpendicular to the *a* and *b* axes, but if the angle should deviate a little from this the length of the *c* axis would be a little greater and would thus approach more closely to Bunn's predicted value of 9·4 Å.

The volume of the cell quoted is 428 Å^3 which is satisfactorily close to that of 440 Å^3 for the β material, and gives a density of 1·05 if two molecular chains are assumed to pass through the cell.

§ 5. DISCUSSION

The examination of β gutta percha has, as was anticipated, given results in general accord with those of previous workers. Consequently, there is little that requires comment except to emphasize the point originally made by Storks (1938) that the orientation of the crystallites perpendicular to the film surface in films whose mean thickness is less than the accepted length of the chain may denote discontinuities within the single molecular chain.

The results are more illuminating in relation to the other modifications of gutta percha leading as they do to the identification of two crystalline forms besides the β modification. The identity periods of 8·76 Å and 9·2 Å are in approximate agreement with the values deduced by Bunn (1942) for two possible alternative configurations for the *trans* polyisoprene chain and thus provide experimental support for the theoretical considerations on which his predictions of the existence of these two forms were based.

The indications of the present work are that, of the two structures under consideration, the γ form is favoured by the unstretched state of the material and the α by the stretched condition. The latter, however, always appears in association with the β form since all stretched samples irrespective of previous heat treatment show a high proportion of β material. The results reported by previous workers when reassessed in the light of this recent evidence also appear to lend support to these conclusions regarding the occurrence of the different forms, and further, indicate that gutta percha as first drawn from the tree exists in the γ form whose characteristic pattern has three strong rings at spacings of 4·95 Å, 3·95 Å, and 3·36 Å and is satisfied by a monoclinic cell with $a=5{\cdot}9$ Å, $b=7{\cdot}9$ Å, $c=9{\cdot}2$ Å, $\gamma=94°$.

The readiness with which this γ structure is broken down either by stretching or heating the material is presumably associated with the geometry of the chain form. The three chain forms for α, β, and γ structures constructed according to the principles laid down by Bunn (1942) are shown diagrammatically in figs. 4 (*a*), (*b*), (*c*). It will be seen from these diagrams that the γ chain differs essentially from the other two forms in two respects: the methyl groups now lie alternately on opposite sides of the chain, and consecutive isoprene units are no longer parallel to one another. The configuration permits compact packing of the chains but does not appear to have the normal fibre characteristics since on stretching the whole of alternate isoprene units swing round the single CH_2-CH_2 bonds connecting them to their neighbours to settle in one of the forms α or β which have the normal fibre characteristics. In his discussion of the behaviour of long chain molecules Bunn remarked that switches from one favoured bond position to another occurred very readily in molecules which contain double bonds within the unit as in the present case. In the present instance the change involves a wide movement of the projecting methyl group and some steric hindrance arising from its passing near to neighbouring molecules might be anticipated, but any such effect is apparently small and readily overcome by the energy put into the system on heating or stretching. It is, however, large enough to render the spontaneous transformation back from the β to the more stable γ form at room temperatures very slow.

It is obvious that some way must be found of obtaining patterns showing much fuller detail before the points discussed can be finally settled. The present work at least appears to have opened up new possibilities and indicates lines along which further research may be prosecuted.

ACKNOWLEDGMENTS

I wish to thank the Council of the Research Association of British Rubber Manufacturers for the grant of a Research Fellowship which made this work possible, and Professor G. I. Finch for his advice and encouragement during its progress. I am also greatly indebted to Mr. C. W. Bunn of Imperial Chemical Industries Ltd., for his most helpful criticism of the work and especially for pointing out the density discrepancy in Bruni and Natta's proposed unit cell for γ gutta percha. Finally I wish to record my thanks to Dr. T. B. Rymer of Reading University for several useful discussions in the later stages of the work, and to the British Rayon Research Association for allowing facilities for its completion.

REFERENCES

BRUNI, G., and NATTA, G., 1934, *Atti Accad. Naz. Lincei*, **19**, 206 (translation: 1934, *Rubber Chem. Tech.*, **7**, 603).

BUNN, C. W., 1942, *Proc. Roy. Soc.* A, **180**, 40, 67, 82.

FINCH, G. I., and FORDHAM, S., 1936, *Proc. Phys. Soc.*, **48**, 85.

FULLER, C. S., 1936, *Industr. Engng. Chem.*, **28**, 907 (and 1937, *Rubber Chem. Tech.*, **10**, 137).

HAUSER, E. A., 1927, *Kautschuk*, **3**, 228.

HAUSER, E. A., and VON SUSICH, G., 1931, *Kautschuk*, **7**, 120, 125, 145.

HOPFF, H., and VON SUSICH, G., 1930, *Kautschuk*, **6**, 234 (translation: 1931, *Rubber Chem. Tech.*, **4**, 75).

MEYER, K. H., and MARK, H., 1930, *Aufbau der hochpolymeren organischen Naturstoffe*, (Leipzig: Akademische Verlagsgesellschaft, M.B.A.), p. 189.

NELSON, J. B., and RILEY, D. P., 1945, *Proc. Phys. Soc.*, **57**, 477.

STILWELL, C. W., and CLARK, G. L., 1931, *Industr. Engng. Chem.*, **23**, 706.

STORKS, K. H., 1938, *J. Amer. Chem. Soc.*, **60**, 1753.

TRZEBIATOWSKI, W., 1937, *Roczn. Chem.*, **17**, 73.

Collisional Processes and Similarity in High-Frequency Discharges in Helium

BY F. LLEWELLYN JONES AND G. C. WILLIAMS
Department of Physics, University College of Swansea

MS. received 19th August 1952

ABSTRACT. The validity of the similarity theorem for high-frequency discharges in helium is examined experimentally in non-uniform fields over a range of frequencies f from 5 to 70 Mc/s and gas pressure p from 1 to 30 mm Hg. The theorem is satisfied for similar discharge systems in pure helium but not when traces of impurity are present. The theorem is shown to be sensitive not only to changes in the parameters pa and f/p, where a is a linear dimension of the discharge tube, but also to changes in the purity of the gas when the impurity can introduce collisional processes which are not dependent on X/p. The failure of the theorem for impure helium for similar discharge systems is accounted for by the occurrence of collisions of the second kind between metastable atoms of helium and molecules of impurity.

§ 1. INTRODUCTION

In a previous paper Llewellyn Jones and Morgan (1951) showed experimentally that the high frequency breakdown potentials V_s for geometrically similar electrode systems in air and hydrogen obey a similarity theorem

$$V_s = \phi(pa, f/p), \qquad \ldots\ldots(1)$$

where p is the gas pressure, a a linear dimension of the tube and f the frequency of the applied field. In this paper the term *similar* high-frequency discharge systems will now be used to refer to geometrically similar systems for which the parameter f/p, as well as the parameter pa, is the same in the two systems.

The basis of the principle of similarity in electrical discharges is that the effective collisional processes should be functions of X/p (Townsend 1915, Holm 1924). Fundamental collisional processes involved in electrical discharges can be regarded as forming two groups: a group which is dependent on the parameter X/p, where X is the electric field, and a group which is not dependent on X/p. Processes of excitation or ionization by single electron impact are typical of the first group since the mean electron energies are dependent on X/p; liberation of electrons from electrodes by field emission (Llewellyn Jones 1949), volume recombination of ions and electrons, and ionization by collisions of the second kind are processes typical of the second group (von Engel and Steenbeck 1934).

Thus, the validity of the similarity theorem for the case of similar discharges in any given gas can give information about the nature of the fundamental collisional processes involved; it has been concluded that the fundamental collisional processes both of electron generation and loss in high-frequency discharges in air (a gas mixture) and in hydrogen (a simple diatomic gas) belong to the group which is dependent on X/p. It remains now to test whether the theorem fails in cases when an essential collisional process in the discharge is not a function of X/p. This paper describes an experimental investigation of the similarity theorem for high-frequency breakdown over a range of oscillation frequencies from 5 to 70 Mc/s in pure helium, and also in helium containing traces of impurity (less than 0·001%) where collisions of the second kind are likely to occur.

§ 2. EXPERIMENTAL PROCEDURE

Three coaxial discharge systems A, B and C were used. These consisted of inner nickel rod electrodes of diameters ($2a$) of 0·159, 0·318 and 0·160 cm respectively, and concentric nickel cylinders of internal diameters ($2b$) of 3·945, 3·950 and 2·080 cm respectively, and the tubes were connected to the same gas system: tubes B and C were geometrically similar. The discharge tubes, the high-frequency oscillator, valve voltmeter, and the general experimental arrangements were the same as those described previously (Llewellyn Jones and Morgan 1951). The gas reservoir contained spectrally pure helium, but before admission to the discharge tubes the gas was stored for at least 15 hours over charcoal cooled by liquid oxygen. To maintain the greatest possible purity of the gas in the discharge tubes, a large glass side-tube containing charcoal in a gauze cage was sealed as near as possible to the tubes and cooled by liquid oxygen (Llewellyn Jones 1931).

The oxide layers formed on the electrodes during exposure to the atmosphere were removed by ion bombardment in discharges in hydrogen maintained in the three discharge tubes simultaneously while the gas was frequently changed. Prolonged ion bombardment of the electrodes finally sputtered the metal on the mica insulating spacers and glass envelope indicating that the treatment was severe.

Then followed successive and alternate periods during which the tubes containing hydrogen were maintained at 350° C in an electric furnace for a period of 10 hours, and were then evacuated for a period of 10 hours while the heating was continued in order to drive off any adsorbed hydrogen. Treatment of this kind had previously been found to be effective for the removal of tenacious oxide layers on nickel (Llewellyn Jones and Davies 1951). When it was considered that all the electrode surfaces were cleaned by this treatment and the discharge tubes had cooled to room temperature, measurements of the breakdown potential were made in helium for the frequency range from 5 to 70 Mc/s and pressures from 1 to 30 mm Hg.

After these measurements the tubes were again maintained at 350° C for some hours and evacuated continuously. At intervals during this treatment the heating was discontinued and the tubes were cooled to room temperature, when fresh specimens of gas were admitted, and breakdown potential measurements made. This procedure was followed for a period of about 50 hours. In this way (V_s, pa) curves were obtained for helium in its final stages of purification. At first the gas contained minute traces of hydrogen liberated from the electrodes during outgassing, but finally the gas was as pure as was possible to maintain it; amounts of impurity present were estimated from the observed very small rate of evolution of gas during outgassing of the tubes. When small traces of impurity were still present, before the outgassing had been completed, it was found that in consecutive measurements of V_s the second determination was higher by approximately 5 V over the whole range of frequencies from 5 to 70 Mc/s. After this all subsequent determinations, when made at intervals of 30 seconds (the time required for a single measurement to be made), were the same to within 0·5 to 1%. This result was taken to indicate that the first discharge drove the final traces of impurities to the walls or to the electrodes, and when the interval between successive discharges was long these impurities were liberated again. This cleaning effect was most marked for high-frequency fields when the concentration of impurities was very low, as found previously by Townsend and MacCallum (1928).

§ 3. EXPERIMENTAL RESULTS

The considerable influence of small traces of impurities on the breakdown potential of helium is shown by the typical set of curves given in fig. 1 relating the breakdown potential V_s and the parameter *pa* for tube C at a frequency of 70 Mc/s. These curves were obtained during the heat treatment described above after removing surface films from the electrodes. The values of V_s for helium not completely purified are given in curve *a*; after some purification by mild heat treatment of the tubes curve *b* was obtained and this differed only slightly from *a*. Continued outgassing, however, increased the values of V_s by approximately 25% to give curve *c*. Further prolonged outgassing at 350° c *in vacuo* produced a further increase in the values of V_s (curve *d*), and the overall effect of heating *in vacuo* for a total of 50 hours was to increase V_s by a factor of 2 from the first measurements. No further increase in V_s with heat treatment was afterwards observed, and a later (V_s, *pa*) curve, *e*, was identical with *d*; the helium was then considered to be pure. Curves of the same general nature were also obtained for the other tubes A and B. A set of (V_s, *pa*) curves for pure helium over the full frequency range is given in fig. 2 for reference, because no previous measurements of the high-frequency breakdown potentials in pure helium in non-uniform field systems are available. It will be seen that V_s decreased as the frequency increased.

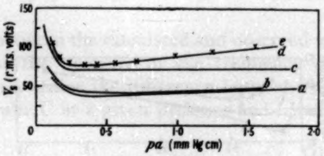


Fig. 1.

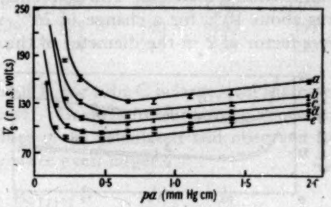


Fig. 2.

Fig. 1. Effect of progressive purification of the gas on the breakdown potential of tube C at 70 Mc/s.

Fig. 2. Similarity relationship in pure helium. The line curves *a*, *b*, *c*, *d*, and *e* are those with tube C at frequencies *f* of 10, 20, 30, 50 and 70 Mc/s, and the points marked are actual measurements obtained with the geometrically similar tube B, of twice the linear dimensions, at half the pressure and at frequencies *f*/2. These points lie on the appropriate curve within the experimental error ; in fact the full curves for B, not given here, coincide under the similarity conditions (equality of *pa* and *f*/*p*) with those of C.

Consider now these results in relation to the similarity theorem. It is necessary first to see whether, for the two geometrically similar tubes B and C, the values of V_s for helium containing traces of impurity were the same over a wide range of *p* and *f*, provided the parameters *pa* and *f*/*p* were the same in each tube. (V_s, *pa*) curves obtained with slightly impure helium are given in fig. 3 for tube B at a frequency of 25 Mc/s (curve *b*), and for tube C at a frequency of 50 Mc/s (curve *a*). These curves should coalesce if the similarity theorem held, but they are, in fact, separate. Consequently, the theorem does not hold for impure helium. Similar results were obtained for any other pair of frequencies from 5 to 70 Mc/s.

Similar results were obtained with impure helium at all stages during the outgassing process when traces of impurities were present. The final measurements of (V_s, *pa*) for tube B at 25 Mc/s and tube C at 50 Mc/s in pure helium, on the other hand, satisfied the theorem as is also shown in fig. 3 (curve *c*).

It is now necessary to test the departures from the similarity theorem in pure helium when one of the parameters *pa* or *f*/*p* was not the same for two tubes. The tubes A and B were not geometrically similar, because the diameters of the wires were in the ratio 1 : 2, but the cylinders were of the same diameter. In this case, it is necessary to compare values of aX_s where

$$aX_s = V_s/\ln(b/a), \qquad \ldots\ldots(2)$$

since the factor $\ln(b/a)$ was no longer the same for both A and B (Llewellyn Jones and Morgan 1951); X_s is the field at the surface of the wire at breakdown. Values of aX_s as a function of *pa* are given in fig. 4 for tube B at a frequency of 25 Mc/s (curve *c*), and for tube A at a frequency of 50 Mc/s (curve *a*). If the similarity theorem still held in this case, even though the tubes were not exactly similar, then the values of aX_s should be the same for a given value of *pa*, and the curves *a* and *c* should coalesce. The difference in the values of aX_s, for A and B, however, was about 10% for a change of 26% in the parameter $\ln(b/a)$ due to a change by a factor of 2 in the diameter of the cylinder.

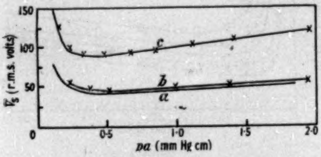


Fig. 3. Similarity theorem for pure and impure helium with tubes B (X) and C (curves).

Curve *a*. C 50 Mc/s, impure helium.
Curve *b*. B 25 ,, ,, ,,
Curve *c*. B 25 ,, C 50 Mc/s, pure helium.

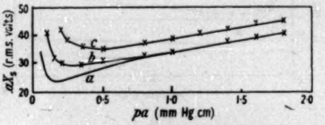


Fig. 4. Effect of variation in the parameters *f*/*p* and ln (*b*/*a*) for pure helium.

Curve *a*. A 50 Mc/s (curve),
Curve *b*. B 50 ,, (X),
Curve *c*. B 25 ,, (X), C 50 Mc/s (curve).

Again, when the parameters *pa* and $\ln(b/a)$ were the same, but the other similarity condition *f*/*p* was not the same—that is for the two tubes B and C—curves *b* and *c* of fig. 4 show that aX_s was different for the two tubes. The values of aX_s for C (curve *c*) are seen to be approximately 10% higher than those for B (curve *b*). The result illustrated was obtained at 50 Mc/s, but similar results were obtained at other frequencies. When both the parameters *pa* and *f*/*p* were the same in each tube the similarity theorem was satisfied.

It has been seen that a variation in the parameter $\ln(b/a)$ of 26% (due to doubling the diameter of the cylinder), or a variation in *f*/*p* by a factor of 2, each produced changes of about 10% in the values of aX_s. When the values of both the parameters each departed from the values corresponding to similarity, then the corresponding changes in aX_s partly cancelled each other ; at 50 Mc/s the cancellation was complete and the values of aX_s were fortuitously the same for tubes A and B over a limited range of *pa* (fig. 4, curves *a* and *b*). The presence of minute traces of hydrogen ($\sim$0·001%), however, had a much greater effect ($\sim$50%) on the magnitudes of the breakdown potentials, so that impurities produced large departures from the similarity theorem. Consequently, experimental measurements of V_s or aX_s can clearly indicate departures in the values of the parameters *pa* and *f*/*p* from the values corresponding to exact similarity in pure gases, and indicate the presence of traces of impurities when the parameters *pa* and *f*/*p* remain constant.

Another important relationship for cylindrical discharges is the corona relationship (Townsend 1915), which relates the breakdown potentials V_1 and V_2 of two discharge systems at the same gas pressure with wires of the same radius but with cylinders of different radii b_1 and b_2 according to the relation

$$V_1 = V_2 \ln(b_1/a)/\ln(b_2/a). \qquad \ldots\ldots(3)$$

In order to apply this relationship to the breakdown potentials of tubes A and C, the values of V_s for tube C, calculated from those measured for tube A using eqn. (3), are compared with the values of V_s actually measured for tube C. Typical data for pure helium are given in the table for extreme values of the pressure p. Even at the highest frequencies and pressures used, the difference between the calculated and observed values of V_s for tube C is large, and increases as the pressure or the frequency decrease. In helium containing traces of impurities the difference between the values of V_s calculated and observed for tube C at a given pressure and frequency were even larger.

p (mm Hg)	V_s (v, r.m.s.) 70 Mc/s		V_s (v, r.m.s.) 50 Mc/s		V_s (v, r.m.s.) 20 Mc/s		V_s (v, r.m.s.) 5 Mc/s	
	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.	Calc.	Obs.
22·5	92	103	105	112	125	140	160	191
2·15	54	88	62	111	84	211	105	223

§ 4. DISCUSSION

The experimental results show that in pure helium the similarity theorem is satisfied well within the experimental accuracy over the full range of frequencies and pressures investigated. In comparing two tubes, small departures in the values of pa and f/p from the similarity condition at once led to inequalities in the corresponding values of V_s. Consequently, the observed differences in the breakdown potentials were clear indications of departures from the similarity conditions in either of the two parameters.

The experimental results show that the processes of ion generation and of loss in pure helium are functions of X/p: these processes are electron generation by single electron collisions with gas atoms and loss by diffusion and drift (Llewellyn Jones and Morgan 1951). The failure of the corona relationship with the discharge tubes used indicates the important part played by electron and ion diffusion and drift in the breakdown mechanism in pure helium. Since the field X_s progressively decreased from the wire to the cylinder, the value of X/p about mid-way between the electrodes was taken and used in the following considerations. Only approximate estimates of the displacement of the electron cloud at lower pressures is possible because of the limited data available on the drift velocities and the diffusion coefficients of electrons in helium for values of X/p greater than 5 v/cm/mm Hg. When V_s for tube C was 112 v at a frequency of 70 Mc/s and a pressure of 25 mm Hg, X and X/p at a radius of 0·5 cm from the rod were 88 v/cm and 3·5 v/cm/mm Hg respectively, and the electron drift velocity was 2×10^7 cm/sec (Healey and Reed 1941). Hence, the time required for the electron cloud to pass from the rod to the cylinder was about ten times greater than the half-period of the field. For an applied potential of 162 v at a frequency of 10 Mc/s when X/p was 5 v/cm/mm Hg the time was of the same order as the half-period of the field. This indicates that the radial drift of the electron cloud, generated near the wire, played an increasing part in the discharge mechanism as the frequency decreased; at the lower frequencies a large fraction of the electrons generated during one half-cycle could have been swept to the electrodes. Similar considerations hold for the loss of electrons by diffusion, and these results indicate the predominating influence of diffusion at the highest frequencies. As the frequency was decreased, however, the influence of the cylinder on both drift and diffusion increased, but at 10 Mc/s the electron loss mechanism was predominantly drift. These results show that the mechanism of the high frequency discharge in pure helium is, therefore, similar to that in air and hydrogen (Llewellyn Jones and Morgan 1951).

In helium containing traces of impurities, on the other hand, the similarity theorem broke down completely; the theorem was not satisfied at any frequency or pressure investigated. This result was obtained in spite of the fact that the helium was spectroscopically pure, and that it was probable that the impurity was less than one part in 10^5. This failure of the similarity theorem for similar discharge systems indicates immediately that the breakdown mechanism in impure helium involved a process, whether of ion generation or loss which was not dependent on X/p. When the impurity was removed, however, the theorem was satisfied and this showed that the process involved was also removed with the impurity. This result can be interpreted by consideration of the fundamental collisional phenomena which occur in high-frequency breakdown (Llewellyn Jones and Morgan 1951).

The impurities, which gave rise to processes not dependent on X/p, had the effect of increasing the rate of generation, or of reducing the rate of loss of ions and electrons, or had both effects. Suppose that the process affected by the impurities was one of electron loss, for example, volume recombination of positive ions and electrons. This process would have acted, however, in addition to drift and diffusion, thus increasing the rate of loss. Consequently, to obtain breakdown the rate of generation would have had to be increased, thus requiring an increase in the breakdown potential. Figure 1 shows that this conclusion is contrary to the experimental observations, which showed that values of V_s for impure helium, when the similarity theorem was not satisfied, were lower than those for pure helium, when the similarity theorem was satisfied. Thus the traces of impurities did not act by introducing a new process of ion loss.

Suppose then that the impurities reduced the loss processes of drift and diffusion by, say, cluster formation, as is known to occur in argon, for example (Llewellyn Jones 1935). Data on monatomic gases not completely pure show that diffusion and drift were always functions of X/p (Townsend 1915, Huxley 1928). Hence, even if, in the present experiments, minute traces of impurities considerably reduced the rate of loss by diffusion and drift, these processes would still remain functions of X/p and would, therefore, still satisfy the similarity theorem. This conclusion is also contrary to the experimental facts which showed that the similarity theorem was not satisfied in impure helium.

The fact that the increased generation in the presence of impurities was accompanied by the failure of the similarity relationship showed that the additional process was not the ionization of molecules of the impurity in single collisions with electrons, because such a process while increasing the ionization would nevertheless still be a function of X/p and would not, therefore, account for the observed failure of the similarity theorem. Calculation of the ratio of the number of molecules of, say, hydrogen impurity ionized to the number of helium atoms excited to the metastable state is not possible because the mean electron energies and their distribution function for values of X/p exceeding about 20 v/cm/mm Hg are not known in helium. However, assuming for simplicity a mean energy of 10 ev, a Maxwellian distribution, and a fractional pressure of hydrogen impurity of 10^{-5}, this ratio is of the order of 10^{-5}, and is negligible.

Suppose, therefore, that the impurities acted by increasing the rate of ionization by some new process of generation not a function of X/p, the new process occurring only when minute traces of impurity were present, while the loss processes remained largely unaffected. This would lead to a reduction in V_s for impure helium compared with the value for pure helium, and would also give values of V_s not satisfying the similarity theorem, because the process introduced was not a function of X/p. These considerations are in accordance with the experimental observations. A possible ionization process is the ionization of molecules of impurity in collisions of the second kind with metastable atoms of helium: the process can be operative when small traces of impurities, of hydrogen for example, are present. This reaction may be expressed by

$$\mathrm{He_{met}} + \mathrm{H_2} \rightarrow \mathrm{He} + \mathrm{H_2}^+ + \mathrm{e}, \qquad \ldots\ldots(4)$$

where $\mathrm{He_{met}}$ represents the metastable atom of helium ($V_{\mathrm{met}} = 19{\cdot}77$ and/or $20{\cdot}4$ ev) and H_2 the normal hydrogen molecule ($V_1^1 = 15{\cdot}6$ ev) provided that $V_{\mathrm{met}} > V_1^1$, a condition which is satisfied for helium and hydrogen, and any of the other gases which may have been adsorbed on the electrodes and glass envelopes. Owing to the long exposure of the electrodes to hydrogen, and the prolonged heat treatment both in hydrogen and *in vacuo*, it was likely that the gas present as an impurity in very small quantities was hydrogen.

Strong support for a reaction of this kind has been obtained by Penning (1931), who showed that in neon containing less than 0·002% of argon the breakdown potential was reduced by at least a factor of 4. Kruithof and his collaborators (Massey and Burhop 1952) have considered the effectiveness of the reaction

$$\mathrm{Ne_{met}} + \mathrm{A} \rightarrow \mathrm{Ne} + \mathrm{A}^+ + \mathrm{e}, \qquad \ldots\ldots(5)$$

and concluded that the chance of this reaction occurring was of the order of unity per collision. Reaction (5) is, however, one of exact resonance, but the energy difference in (4) is quite large and its efficiency correspondingly lower. However, if allowance is made for the dissociation of H_2 (dissociation energy $\sim 4{\cdot}5$ ev) the total excitation energy can be transferred in a collision. In this case atomic ions are formed.

These results show that the similarity theorem may now be used as a tool to investigate the relative importance of various collisional phenomena in discharges, in particular, the group of important processes which are not functions of the parameter X/p.

ACKNOWLEDGMENTS

We wish to thank the Department of Scientific and Industrial Research for the award of a grant, and also the Council of the University of Wales for the award of a Fellowship to one of us (G. C. W.) to undertake this work.

REFERENCES

von Engel, A., and Steenbeck, M., 1934, *Elektrische Gasentladungen*, Vol. II (Berlin: Springer) p. 95.

Healey, R. H., and Reed, J. W., 1941, *The Behaviour of Slow Electrons in Gases* (Sydney: Amalgamated Wireless (Australia) Ltd.).

Holm, R., 1924, *Phys. Z.*, **25**, 497.

Huxley, L. G. H., 1928, *Phil. Mag.*, **5**, 721.

Llewellyn Jones, F., 1931, *Phil. Mag.*, **11**, 163; 1935, *Proc. Phys. Soc.*, **47**, 74; 1949, *Ibid.*, B, **62**, 366.

Llewellyn Jones, F., and Davies, D. E., 1951, *Proc. Phys. Soc.* B, **64**, 519.

Llewellyn Jones, F., and Morgan, G. D., 1951, *Proc. Phys. Soc.* B, **64**, 560.

Massey, H. S. W., and Burhop, E. H. S., 1952, *Electronic and Ionic Impact Phenomena* (Oxford: Clarendon Press).

Penning, F. M., 1931, *Phil. Mag.*, **11**, 961.

Townsend, J. S., 1915, *Electricity in Gases* (Oxford: Clarendon Press).

Townsend, J. S., and MacCallum, S. P., 1928, *Phil. Mag.*, **6**, 857.

Perturbations in the Magnetic Deflector for Synchro-Cyclotrons

By K. J. LE COUTEUR
Department of Theoretical Physics, University of Liverpool

Communicated by H. Fröhlich; *MS. received* 8*th May* 1952, *and in final form* 1*st October* 1952

ABSTRACT. Previously the linearized equation of motion of particles in the deflector were solved exactly; now the effect of non-linear terms, proportional to d^2H/dr^2, is evaluated approximately. The results indicate the tolerances on the field shape and the desirable radial position of the deflector.

---

§ 1. INTRODUCTION

Tuck and Teng (1950) proposed to extract the beam from a synchro-cyclotron by modifying the normal magnetic field in such a way as to build up the amplitude of radial oscillation until protons can escape from the magnet. Le Couteur (1951) studied the behaviour of the deflector analytically and determined operating conditions which should provide adequate radial deflection without leading to vertical instability of the beam. For mathematical reasons this work assumed a linear variation of H with radius over the region of deflection, but within this assumption the analysis was exact. This condition is not realized precisely in practice, and the purpose of the present paper is to examine the effect of a non-linear radial variation of H upon the performance of the deflector.

Non-linearity introduces terms, proportional to d^2H/dr^2, which couple the radial and vertical motions of the beam. In normal synchro-cyclotron operation the effects of coupling vanish when averaged over a few revolutions because the radial and vertical oscillations are uncorrelated: an exception occurs at the radius $n=0{\cdot}2$ where the frequencies of the two come into resonance; then the beam usually hits the dee because of the transfer of energy from radial to vertical motion (Heinrich, Sewell and Vale 1949). The peeler–regenerator directs the particle into a radial motion of fixed phase and increasing amplitude and in so doing necessarily influences the vertical motion; there results some correlation between the two motions which allows the effect of their mutual coupling to build up from turn to turn. If the correlation and the coupling coefficient d^2H/dr^2 are sufficiently large a vertical blow up may occur; the effect is similar to that usually associated with $n=0{\cdot}2$ but, because of the correlation, in the deflector it may occur at an earlier radius.

Since these difficulties do not arise if d^2H/dr^2 vanishes they can obviously be avoided by placing the deflector well inside the synchro-cyclotron magnet so that the variation of dH/dr over the region of deflection is negligible; the requirements are evaluated quantitatively in § 4. This solution has the disadvantage that deflection must start several inches inside the $n=0{\cdot}2$ radius, so that particle acceleration cannot continue to the full energy of which the machine would otherwise be capable. To extract the greatest possible energy the deflector must be placed close to the edge of the synchro-cyclotron, and then the rapid fall off of the magnetic field supplies part of the required peeler effect. The analysis of § 5 suggests that stability against vertical blow up can be maintained by proper disposition of the regenerator.

§ 2. EQUATIONS OF MOTION

In the previous paper on the regenerative deflector (Le Couteur 1951, to be referred to as I)* the linearized particle equations of motion were solved exactly. This procedure is valid if d^2H/dr^2 is sufficiently small, where H is the vertical component of magnetic field strength.

If this condition is not satisfied equations (1) of I should be replaced by the second-order equations of motion :

$$\frac{1}{\nu^2}\frac{d^2\rho}{dt^2}+\left(1+\frac{r}{H}\frac{\partial H}{\partial r}\right)\rho=\frac{1}{2}\frac{r}{H}\frac{\partial^2 H}{\partial r^2}(z^2-\rho^2)+\frac{1}{2r}\left(3+\frac{r}{H}\frac{\partial H}{\partial r}\right)\rho^2, \qquad \ldots\ldots(1\,a)$$

$$\frac{1}{\nu^2}\frac{d^2z}{dt^2}-\frac{r}{H}\frac{\partial H}{\partial r}z=\frac{r}{H}\frac{\partial^2 H}{dr^2}\rho z, \qquad \ldots\ldots(1\,b)$$

where $\nu \simeq d\theta/dt$ is the synchronous frequency and all coefficients are to be evaluated on the synchronous orbit $r=r_s$. These equations may be derived from the relativistic Hamiltonian (see Appendix).

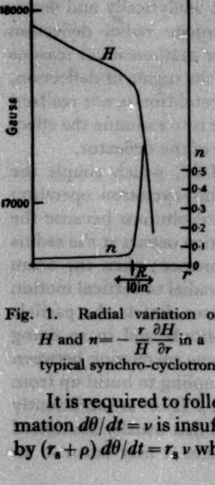


Fig. 1. Radial variation of H and $n=-\frac{r}{H}\frac{\partial H}{\partial r}$ in a typical synchro-cyclotron.

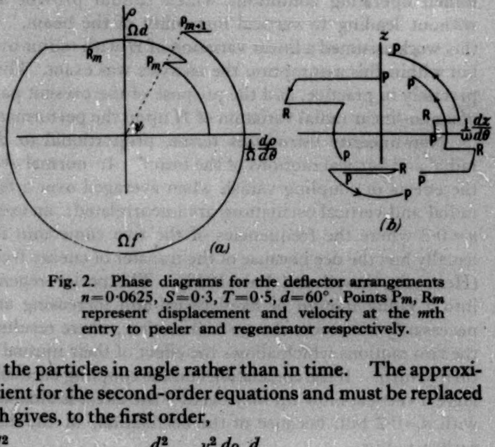

ρ
Ωd
P_{m+1}
R_m
P_m
ψ
$\frac{1}{\Omega}\frac{d\rho}{d\theta}$
Ωf
(a)
z
P
R
$\frac{1}{\omega}\frac{dz}{d\theta}$
(b)

Fig. 2. Phase diagrams for the deflector arrangements $n=0{\cdot}0625$, $S=0{\cdot}3$, $T=0{\cdot}5$, $d=60°$. Points P_m, R_m represent displacement and velocity at the mth entry to peeler and regenerator respectively.

It is required to follow the particles in angle rather than in time. The approximation $d\theta/dt=\nu$ is insufficient for the second-order equations and must be replaced by $(r_s+\rho)\,d\theta/dt=r_s\,\nu$ which gives, to the first order,

$$\frac{d^2}{dt^2}=\nu^2(1-2\rho/r_s)\frac{d^2}{d\theta^2}-\frac{\nu^2}{r_s}\frac{d\rho}{d\theta}\frac{d}{d\theta}. \qquad \ldots\ldots(2)$$

Substitution in (1) yields the equations of motion

$$\frac{d^2\rho}{d\theta^2}+\left(1+\frac{r}{H}\frac{\partial H}{\partial r}\right)\rho=\frac{1}{2}\frac{r}{H}\frac{\partial^2 H}{\partial r^2}(z^2-\rho^2)+\frac{1}{r}\left(\frac{d\rho}{d\theta}\right)^2-\frac{1}{2r}\left(1-\frac{r}{H}\frac{\partial H}{\partial r}\right)\rho^2 \qquad ..(3\,a)$$

$$\frac{d^2z}{d\theta^2}-\frac{r}{H}\frac{\partial H}{\partial r}z=\frac{r}{H}\frac{\partial^2 H}{\partial r^2}\rho z+\frac{2}{H}\frac{\partial H}{\partial r}\rho z+\frac{1}{r}\frac{d\rho}{d\theta}\frac{dz}{d\theta}, \qquad \ldots\ldots(3\,b)$$

---

* *Erratum.* The bottom right-hand element of the matrices P and Q in equations (A1), (B1), (C6), (C9) of I should read 1, not 0.

and on the right-hand side the terms proportional to d^2H/dr^2 are usually the most important.

In actual synchro-cyclotrons d^2H/dr^2 is negligible near the centre of the magnet but not at the edge, where the field strength falls off suddenly as shown in fig. 1. The radius at which d^2H/dr^2 increases suddenly from a small to a large value is denoted by R.

Deflection of the particles starts at a radius r_s and leads to radial oscillations about r_s, which build up to an amplitude of 3 or 4 inches. We are therefore concerned with the particle equations of motion over the region $r_s \pm 3$ in. of the magnet. In this paper eqns. (1) will be studied in the two important limiting cases (*a*) $r_s < R$ so that in the region $r_s \pm 3$ in. of deflection d^2H/dr^2 may be treated as constant and small; (*b*) $r_s \simeq R$ so that one may assume d^2H/dr^2 is constant for positive values of ρ and negligible for negative values. The main problem is the effect of the coupling terms upon the stability of the vertical motion when the radial oscillation is large.

§ 3. THE SOLUTION WHEN $d^2H/dr^2=0$

The radial displacement at angle θ has the form (I, eqn. (13))

$$\rho = ke^{m\Lambda} \sin(\psi + \Omega\theta) \qquad \ldots\ldots(4)$$

at the *m*th passage through gap *d*, and

$$\rho = ke^{m\Lambda} \frac{\sin(\psi+\Omega d)}{\sin\phi} \sin(\phi - \Omega d + \Omega\theta) \qquad \ldots\ldots(5)$$

at the *m*th passage through gap *f*. This is illustrated in fig. 2 (*a*) by a phase diagram. In such diagrams the peeler and regenerator contribute impulses

$$\left(\frac{d\rho}{d\theta}\right)_{\text{out}} - \left(\frac{d\rho}{d\theta}\right)_{\text{in}} = S\rho \quad \text{and} \quad \left(\frac{d\rho}{d\theta}\right)_{\text{out}} - \left(\frac{d\rho}{d\theta}\right)_{\text{in}} = -T\rho \qquad \ldots\ldots(6)$$

to the radial motion, therefore the phase ϕ in f is given by

$$\cot\phi = \cot(\psi + \Omega d) - T/\Omega. \qquad \ldots\ldots(7)$$

The displacement ρ vanishes at the points in gap f where $\theta = d + (\pi - \phi)/\Omega$ and $\theta = d + (2\pi - \phi)/\Omega$.

The amplitude of vertical oscillation fluctuates about a constant value; the motion is shown in fig. 2 (*b*) for several values of the varying phase of entry into the peeler.

§4. THE SOLUTION WHEN d^2H/dr^2 IS SMALL AND CONSTANT (CASE (*a*))

In the gaps between peeler and regenerator, eqn. (3 *b*) has an energy integral

$$[\omega^2h^2]_{\theta_1}^{\theta_2} \equiv \left[\omega^2 z^2 + \left(\frac{dz}{d\theta}\right)^2\right]_{\theta_1}^{\theta_2} = 2\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right)\int_{\theta_1}^{\theta_2} \rho z \frac{dz}{d\theta}\, d\theta + \frac{2}{r}\int_{\theta_1}^{\theta_2} \frac{d\rho}{d\theta}\left(\frac{dz}{d\theta}\right)^2 d\theta, \qquad \ldots\ldots(8)$$

where $\omega^2 = n = -(r/H)(\partial H/\partial r)$ and h is the amplitude of vertical oscillation.

In normal synchro-cyclotron operation there is no net transfer of energy from the radial to the vertical oscillation because when a few revolutions are considered the right-hand side of (3) averages to zero; there is, however, a well-known exception if the frequencies Ω and ω of radial and vertical oscillation satisfy the resonance condition $\Omega = 2\omega$, which occurs at $n = 0{\cdot}2$.

In the regenerative deflector the radial motion has the well-defined phase shown in fig. 2 and although the phase of the vertical oscillation is not fixed the peeler and regenerator produce some coherence of the vertical motion so that at any given angular position the average value of $z\,dz/d\theta$ is not zero. The first integral on the right-hand side of (8) therefore grows with each orbital revolution.

With the above assumption (*a*) the integrals in (8) may be calculated for the unperturbed motion described in § 3, and the second integral averages to zero.

Let $\bar{h}$ denote the average amplitude of vertical oscillation. Then for the numerical values used in fig. 2 the first integral (8) over gaps d and f fluctuates according to the phase between $-1{\cdot}5\ \omega\bar{h}^2ke^{m\Lambda}$ and $-0{\cdot}7\ \omega\bar{h}^2ke^{m\Lambda}$ for the mth revolution; the average value $-1{\cdot}1\ \omega\bar{h}^2ke^{m\Lambda}$ is made up of contributions $-0{\cdot}45\ \omega\bar{h}^2ke^{m\Lambda}$ from gap d and $-0{\cdot}65\ \omega\bar{h}^2ke^{m\Lambda}$ from gap f.

If the radial variation of the peeler field is non-linear there is a further contribution

$$\theta_p\left[\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2}+\frac{2}{H}\frac{\partial H}{\partial r}\right)\rho z\left\{\left(\frac{dz}{d\theta}\right)_{\text{in}}+\left(\frac{dz}{d\theta}\right)_{\text{out}}\right\}\right]_{\text{peeler}}$$

$$=-\left(\frac{dS}{dr}+\frac{S}{r}\right)\left[\rho z\left\{\left(\frac{dz}{d\theta}\right)_{\text{in}}+\left(\frac{dz}{d\theta}\right)_{\text{out}}\right\}\right]_{\text{peeler}} \qquad \ldots\ldots(9)$$

to the right-hand side of (8) and a similar contribution

$$\left(\frac{dT}{dr}+\frac{T}{R}\right)\left[\rho z\left\{\left(\frac{dz}{d\theta}\right)_{\text{in}}+\left(\frac{dz}{d\theta}\right)_{\text{out}}\right\}\right]_{\text{regen}} \qquad \ldots\ldots(10)$$

from the regenerator. In the above example (9) fluctuates between $1{\cdot}4$ and $-0{\cdot}9$ times $\omega\bar{h}^2ke^{m\Lambda}(dS/dr+S/r)$ with an average value of $0{\cdot}45\ \omega\bar{h}^2ke^{m\Lambda}(dS/dr+S/r)$ and (10) fluctuates between $0{\cdot}6$ and $-0{\cdot}9$ times $\omega\bar{h}^2ke^{m\Lambda}(dT/dr+T/r)$, with average value approximately zero.

The average increase in amplitude of vertical oscillation from the first to the final revolution is thus

$$[h^2]_0^{\mathrm{f}}=\frac{\bar{h}^2}{\omega}\left[-2{\cdot}2\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2}+\frac{2}{H}\frac{\partial H}{\partial r}\right)+0{\cdot}45\left(\frac{dS}{dr}+\frac{S}{r}\right)\right]\int_0^{\mathrm{f}} ke^{m\Lambda}\,dm$$

$$=\frac{\bar{h}^2}{\omega\Lambda}\left[2{\cdot}2\left(\frac{\partial n}{\partial r}+\frac{n}{r}\right)+0{\cdot}45\left(\frac{dS}{dr}+\frac{S}{r}\right)\right]\rho_{\mathrm{f}} \qquad \ldots\ldots(11)$$

where ρ_{f} is the maximum radial displacement at the last revolution, that is, the radial distance from r_s to the magnetic channel. Both terms in the bracket of eqn. (11) are necessarily positive.

The amplitude increase (11) is mainly determined by the magnitude of the coupling coefficients on the right-hand side of (3 *b*) and by the extent of the radial displacement; the factor $1/\Lambda$ represents the number of revolutions a particle makes in the deflector before extraction, and the factor $1/\omega$ represents the effect of the first-order vertical focusing forces in restraining the increase of amplitude.

For satisfactory operation of the deflector the increase of h must be limited and a reasonable restriction is $h_{\mathrm{f}}^2-h_0^2<\frac{1}{2}\bar{h}^2$, which requires

$$\left[2{\cdot}2\left(\frac{\partial n}{\partial r}+\frac{n}{r}\right)+0{\cdot}45\left(\frac{\partial S}{\partial r}+\frac{S}{r}\right)\right]\rho_{\mathrm{f}}<\tfrac{1}{2}\omega\Lambda,$$

but as the derivatives are difficult to measure this is more conveniently expressed as*

$$2{\cdot}2\Delta n + 0{\cdot}45\Delta S + \left(2{\cdot}2\frac{n}{r} + 0{\cdot}45\frac{S}{r}\right)\rho_{\rm f} < \tfrac{1}{2}\omega\Lambda, \qquad \ldots\ldots(12)$$

in which Δn and ΔS are the increases of n and S over the displacement $\rho_{\rm f}$ from $r_{\rm s}$ to the magnetic channel. To limit the fluctuating peeler and regenerator contributions (9) and (10) additional restrictions

$$|\Delta T| = |dT/dr|\rho_{\rm f} < \tfrac{1}{2}\omega \quad \text{and} \quad |\Delta S| = |dS/dr|\rho_{\rm f} < \tfrac{1}{2}\omega \qquad \ldots\ldots(13)$$

should be imposed.

The additional terms on the right-hand side of (3 *a*) produce a negligible displacement of the centre of the radial oscillations.

In a typical case one has $\omega = \frac{1}{4}$, $\Lambda = 0{\cdot}4$, so (12) requires $\Delta n < 0{\cdot}02$. Reference to fig. 1 shows that in a typical case this restriction can be satisfied if the magnetic channel is placed at radius R and deflection starts at radius $r_{\rm s} = R - 3$ in., for n increases little between these radii. The disadvantage of this arrangement is that the energy of the extracted beam corresponds to radius $R - 3$ in., so that the further acceleration which might be achieved by using the last few inches of the synchrocyclotron is sacrificed.

§ 5. THE SOLUTION WHEN $r_{\rm s} \simeq R$ (CASE (*b*))

Figure 2 shows that the regions where ρ is positive and d^2H/dr^2 large occur near the peeler and regenerator. The effect on the vertical motion of the terms on the right-hand side of (3 *b*) may be represented by impulses

$$\int_{\varrho=0}^{\theta=d/2}\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right)\rho z\,d\theta \quad \text{and} \quad \int_{\theta=d/2}^{\varrho=0}\left(\frac{r}{H}\frac{\partial^2 H}{dr^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right)\rho z\,d\theta$$

localized at peeler and regenerator respectively. These are equivalent to replacing the peeler and regenerator strengths S and T by

$$S_{\rm v} = S - \int_{\varrho=0}^{\theta=d/2}\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right)\rho\,d\theta \quad \text{and} \quad T_{\rm v} = T + \int_{\theta=d/2}^{\varrho=0}\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right)\rho\,d\theta,$$

which are effective strengths for the vertical motion. The integrals are taken between the points at which ρ becomes positive and the point $\theta = \frac{1}{2}d$ midway between peeler and regenerator. The numerical values used in fig. 2 lead to

$$\left.\begin{aligned} S_{\rm v} &= S - 0{\cdot}91ke^{m\Lambda}\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right) \simeq S + 0{\cdot}91\,\Delta n, \\ T_{\rm v} &= T + 0{\cdot}92ke^{m\Lambda}\left(\frac{r}{H}\frac{\partial^2 H}{\partial r^2} + \frac{2}{H}\frac{\partial H}{\partial r}\right) \simeq T - 0{\cdot}92\,\Delta n, \end{aligned}\right\} \qquad \ldots\ldots(14)$$

where $\Delta n_{\rm s}$ is the increase of n between $r_{\rm s}$ and the radius $r_{\rm s} + k\,e^{m\Lambda}$ which is reached at the mth revolution. If $r_{\rm s}$ is close to R, at the final revolution Δn may be as great as $0{\cdot}2$. Such a large change in the effective peeler and regenerator strengths would probably carry the values $S_{\rm v}$, $T_{\rm v}$ outside the region of vertical stability in fig. 4 of I or at least to regions where the amplitude of vertical oscillation is very large. It may, however, be possible to avoid vertical instability by choosing a weak peeler S and strong regenerator T, and by curving the regenerator field so that T increases outwards. Such an arrangement could keep the effective strengths $S_{\rm v}$, $T_{\rm v}$ within the requirements of stability.

---

* Strictly, the coefficients 2·2 and 0·45 refer only to the example shown in fig. 2.

The effect on the radial motion of the large terms on the right-hand side of (3 *a*) can similarly be represented by impulses

$$-\frac{1}{2}\int \frac{r}{H}\frac{\partial^2 H}{\partial r^2}\rho^2\,d\theta$$

if z^2 is neglected in comparison with ρ^2. These are equivalent to replacing S and T by

$$S_R = S - \frac{1}{2}\int \frac{r}{H}\frac{\partial^2 H}{\partial r^2}\frac{\rho^2\,d\theta}{\rho_{\text{peeler}}} \quad\text{and}\quad T_R = T + \frac{1}{2}\int \frac{r}{H}\frac{\partial^2 H}{\partial r^2}\frac{\rho^2\,d\theta}{\rho_{\text{regen}}},$$

which are effective strengths for the radial motion. The numerical values of fig. 2 lead to

$$\left.\begin{aligned} S_R &= S - 0{\cdot}41\frac{r}{H}\frac{\partial^2 H}{\partial r^2}ke^{m\Delta} \simeq S + 0{\cdot}41\Delta n,\\ T_R &= T + 0{\cdot}40\frac{r}{H}\frac{\partial^2 H}{\partial r^2}ke^{m\Delta} \simeq T - 0{\cdot}40\Delta n \end{aligned}\right\} \qquad \ldots\ldots(15)$$

and this change would slightly decrease the radial gain.

As an example, suppose that n increases from 0·06 to 0·26 over 3 in. outwards from r_s, and that peeler–regenerator strengths $S = 0{\cdot}2$, $T = 0{\cdot}7$, are chosen. As the increasing radial oscillation carries the particles into the fringing field of the magnet, (14) shows that the vertical motion gradually becomes characterized by effective strengths $S = 0{\cdot}4$, $T = 0{\cdot}5$, and the radial motion by effective strengths $S = 0{\cdot}3$, $T = 0{\cdot}6$. If the peeler–regenerator angle is $d = 60°$, fig. 4 (*b*) of I shows that these variations leave the radial gain almost unchanged at $e^{\Lambda} = 1{\cdot}5$ and that vertical focusing is maintained. In the final stages of deflection most of the peeler action is provided by the fall off of the normal magnetic field. The energy of the extracted beam should be that achieved at radius $r_s \simeq R$, which might be 5% more than that achieved by the arrangement of § 4.

§ 6. CONCLUSIONS

In arrangement (*a*) deflection starts at a smaller radius than in (*b*): the former is less critical and should yield a higher beam intensity, but perhaps 5% less particle energy than the latter. The necessary restrictions on the curvature of the magnetic fields are given by eqns. (12), (13), (14), (15), and are discussed at the ends of § 4 and § 5. Since the curvature increases rapidly towards the edge of the magnet, these conditions in effect determine the desirable radial position of the deflector.

ACKNOWLEDGMENT

I am indebted to Mr. H. Newns for checking eqns. (1).

APPENDIX

THE SECOND-ORDER EQUATIONS OF MOTION

As in § 2 of I, the magnetic field is represented by the vector potential A and, following Corben and Stehle (1950), eqn. (104.8), the relativistic Hamiltonian for the motion of the particle is

$$\mathscr{H} \equiv c\{m^2c^2 + p_r^{\,2} + p_z^{\,2} + (p_\theta/r - eA/c)^2\}^{1/2} = \tilde{m}c^2 \qquad \ldots\ldots(\text{A}\,1)$$

where $\tilde{m}$ is the apparent mass and the generalized angular momentum $p_\theta = \tilde{m}r^2\,d\theta/dt + erA/c$ is a constant of the motion. It is convenient to write

$$P = p_\theta/r - eA/c \qquad \ldots\ldots(\text{A}\,2)$$

for the tangential momentum $\tilde{m}r\,d\theta/dt$, so that

$$\mathscr{H}=c\{m^2c^2+p_r{}^2+p_z{}^2+P^2\}^{1/2} \qquad \ldots\ldots\text{(A 3)}$$

obviously represents the kinetic energy of the particle. Then

$$\frac{1}{c^2}\frac{\partial\mathscr{H}}{\partial z}=\frac{P}{\mathscr{H}}\frac{\partial P}{\partial z}=-\frac{P}{\mathscr{H}}\frac{e}{c}\frac{\partial A}{\partial z} \qquad \ldots\ldots\text{(A 4)}$$

and
$$\frac{1}{c^2}\frac{\partial\mathscr{H}}{\partial r}=\frac{P}{\mathscr{H}}\frac{\partial P}{\partial r}=-\frac{P}{\mathscr{H}}\left(\frac{p_\theta}{r^2}+\frac{e}{c}\frac{\partial A}{\partial r}\right) \qquad \ldots\ldots\text{(A 5)}$$

must vanish on the synchronous orbit, which is thus determined for a particle of momentum P as

$$z=0 \quad\text{and}\quad P=eH_zr/c,\ d\theta/dt=eH_z/\tilde{m}c=\nu \qquad \ldots\ldots\text{(A 6)}$$

where
$$H_z=-A/r-\partial A/\partial r \text{ and } H_r=\partial A/\partial z \qquad \ldots\ldots\text{(A 7)}$$

are the magnetic field strengths (following Corben–Stehle, left-hand axes are used).

For small oscillations about this orbit one must evaluate the derivatives (A 4), (A 5) at points displaced by ρ and z from the synchronous orbit. These may be expressed in terms of higher derivatives, *which are evaluated on the synchronous orbit*; many terms vanish because of the symmetry about $z=0$. Thus (to the second order)

$$\frac{\partial\mathscr{H}}{\partial z}(\rho,z)=z\frac{\partial^2\mathscr{H}}{\partial z^2}+\rho z\frac{\partial^3\mathscr{H}}{\partial r\partial z^2}=\frac{c^2P}{\mathscr{H}}\left(z\frac{\partial^2P}{\partial z^2}+\rho z\frac{\partial^3P}{\partial r\partial z^2}\right) \qquad \ldots\ldots\text{(A 8)}$$

and
$$\frac{\partial\mathscr{H}}{\partial r}(\rho,z)=\rho\frac{\partial^2\mathscr{H}}{\partial r^2}+\tfrac{1}{2}\rho^2\frac{\partial^3\mathscr{H}}{\partial r^3}+\tfrac{1}{2}z^2\frac{\partial^3\mathscr{H}}{\partial r\partial z^2}$$

$$=\frac{c^2P}{\mathscr{H}}\left(\rho\frac{d^2P}{dr^2}+\tfrac{1}{2}\rho^2\frac{d^3P}{dr^3}+\tfrac{1}{2}z^2\frac{d^3P}{dr\,dz^2}\right). \qquad \ldots\ldots\text{(A 9)}$$

The derivatives are easily calculated from (A 2) and expressed in terms of H_z and of the synchronous frequency ν. Then the canonical equations of motion

$$\tilde{m}\frac{d^2z}{dt^2}=\frac{dp_z}{dt}=-\frac{\partial\mathscr{H}}{\partial z},\quad \tilde{m}\frac{d^2\rho}{dt^2}=\frac{dp_r}{dt}=-\frac{\partial\mathscr{H}}{\partial r} \qquad \ldots\ldots\text{(A 10)}$$

reduce to (1 *a*) and (1 *b*); for example,

$$\frac{d^2z}{dt^2}=\frac{ceP}{\tilde{m}\mathscr{H}}\left(z\frac{\partial^2A}{\partial z^2}+\rho z\frac{\partial^3A}{\partial r\partial z^2}\right)=\frac{er\nu}{\tilde{m}c}\left(z\frac{\partial H_r}{\partial z}+\rho z\frac{\partial^2H_r}{\partial r\partial z}\right)$$

$$=\nu^2\frac{r}{H_z}\left(z\frac{\partial H_z}{\partial r}+\rho z\frac{\partial^2H_z}{\partial r^2}\right).$$

Alternatively, one can work from the Lorentz equations of motion, and the derivatives of H_z and H_r appear to represent the variation of the Lorentz force as the particle is displaced from the synchronous orbit. The second-order terms in the radial Lorentz force involve the first-order change in azimuthal velocity and eqn. (1 *a*) results if this is determined from the integral of motion $p_\theta=$ constant. If instead $d\theta/dt$ is expressed in terms of $d\rho/dt$ and dz/dt by means of the energy equation, the second-order equations of motion become much more complicated (Hamilton and Lipkin 1951, Appendix V).

The small terms in (1 *a*) may be checked by considering the special case of constant H_z. Equation (1 *a*) is satisfied by $r=(a^2+l^2+2al\cos\nu t)^{1/2}$, $\rho=r-a+\frac{1}{2}l^2/a$ which represents motion on a circle of radius *a* with centre displaced to a distance *l* from the origin.

REFERENCES

Corben, H. C., and Stehle, P., 1950, *Classical Mechanics* (New York: Wiley).

Hamilton, D. R., and Lipkin, H. J., 1951, *Rev. Sci. Instrum.*, **22**, 783.

Heinrich, L. R., Sewell, D. C., and Vale, J., 1949, *Rev. Sci. Instrum.*, **20**, 887.

Le Couteur, K. J., 1951, *Proc. Phys Soc.* B, **64**, 1073.

Tuck, J. L., and Teng, L. C., 1950, *Institute of Nuclear Studies*, 170 in. *Synchro-cyclotron Progress Report III* (University of Chicago), Chap. VIII.

Spectroscopic Study of Caesium Discharges in a Magnetic Field

By L. W. DAVIES
Clarendon Laboratory, Oxford

Communicated by H. G. Kuhn; *MS. received* 18*th August* 1952

ABSTRACT. From measurements of the intensity distribution in the recombination spectrum, the relative densities of the electrons and their velocity distribution in the positive column of a caesium discharge are determined in the presence of a longitudinal magnetic field. In radio-frequency discharges, the magnetic field is found to increase the electron temperature, contrary to expectation. In d.c. discharges, the effect of the field is found to depend on the direction of the current; this effect is ascribed to gradients in the gas pressure set up by the discharge current.

---

§ 1. INTRODUCTION

The experiments described in this paper consist of spectrographic determinations of the distribution of intensity in the recombination radiation emitted by the positive column of a discharge in caesium vapour. Discharges were investigated under various conditions, particularly in the presence of a longitudinal magnetic field. The intensity distribution in the 6P recombination continuum of caesium has been related to the distribution of electron velocities in the discharge (Mohler and Boeckner 1929, Boeckner 1931, Mohler 1936, 1938 a): if $J(\nu)$ denotes the intensity of the 6P recombination radiation of frequency ν, then

$$J(\nu)=\text{const. } N_1(c)N_2\nu^{-1}(\nu-\nu_{6\mathrm{P}})^{-1}, \qquad \ldots\ldots(1)$$

where $N_1(c)\,dc$ is the number of electrons per unit volume with speeds in the range $(c, c+dc)$, N_2 is the density of positive ions and $\nu_{6\mathrm{P}}$ is the frequency of the 6P series limit. Spectrographic determinations of $J(\nu)$ consequently enabled the distribution $N_1(c)$ to be calculated in each case.

In the positive column of a gas discharge in which there are no negative ions we may assume that $N_1=N_2$, where N_1 is the electron density; if in addition we have a Maxwellian distribution of electron velocities, eqn. (1) becomes

$$J(\nu)=\text{const. } N_1{}^2T_1{}^{-3/2}\nu^{-1}\exp(-\boldsymbol{h}\nu/\boldsymbol{k}T_1), \qquad \ldots\ldots(2)$$

where T_1 is the electron temperature. This yields

$$\ln\{\nu J(\nu)\}=-\boldsymbol{h}\nu/\boldsymbol{k}T_1+\text{const.} \qquad \ldots\ldots(3)$$

A linear relationship between experimental values of $\log\{\nu J(\nu)\}$ and frequency ν thus indicates the existence of a Maxwellian distribution of electron velocities, and the value of T_1 is readily calculable from the spectrographic measurements.

In all the experiments described here it was found that there was a Maxwellian distribution of electron velocities, within the experimental error. In this case there is a simple relationship which enables one to compare the electron densities at the corresponding points of two discharges: if subscripts *a* and *b* are used to denote the parameters of two discharges, eqn. (2) for the two cases gives

$$N_{1a}/N_{1b}=\{J_a(\nu)/J_b(\nu)\}^{1/2}(T_{1a}/T_{1b})^{3/4}\exp[\boldsymbol{h}\nu(T_{1a}^{-1}-T_{1b}^{-1})/2\boldsymbol{k}]. \quad \ldots\ldots(4)$$

Rewriting (4), we may also derive

$$\frac{d}{d\nu}[\log\{J_a(\nu)/J_b(\nu)\}]=-2\cdot084\times10^{-11}(T_{1a}^{-1}-T_{1b}^{-1}). \quad \ldots\ldots(5)$$

This relationship allows an accurate determination of the change produced in the value of the electron temperature by a particular variation in the discharge conditions (see § 2).

It has been suggested (Jancke 1936, Finkelnburg 1938, 1948) that a hypothetical process, ‘preferential recombination’, also gives rise to continuous radiation on the short-wavelength side of the series limit in atomic spectra. This process was suggested in order to explain a correlation observed between the intensities of the continuum and of the higher series lines lying to the long-wavelength side of the series limit; since such a correlation has been satisfactorily explained (Giovanelli 1948) in terms of processes other than ‘preferential recombination’, the latter may safely be neglected. Boeckner (1931) and Mohler (1936, 1937) have found additional experimental and theoretical evidence in favour of the interpretation of the continuum as recombination radiation.

§ 2. APPARATUS

The discharge tubes used were of Pyrex glass (fig. 1); the positive column section was 10 cm long and 5 mm in internal diameter. Each of the two electrodes was constructed of thin nickel sheet, and was capable of thermionic emission of up to 2·5 A; the tube was operated both with d.c., with either electrode acting as cathode, and with radio-frequency excitation. The sealed-off discharge tube was heated in an oven during the experiments, in order to obtain the required value of the vapour pressure; the liquid caesium condensed in the appendix tube, where its temperature was measured and the corresponding vapour pressure calculated from a formula given by Ditchburn and Gilmour (1941). Spectrographic examination of the radiation from the discharges indicated that the only impurities present were traces of other alkali metals.

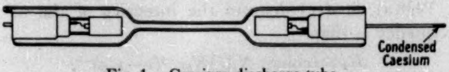


Fig. 1. Caesium discharge tube.

The longitudinal magnetic field was provided by a horizontal single-layer solenoid of 27 turns, its length being 23 cm; currents of up to 1000 A could be passed continuously through the solenoid, giving a magnetic field along the positive column of calculated intensity 1580 gauss. The positive column was aligned centrally, with its axis coincident with that of the solenoid.

An image of the central portion of the positive column was focused on the spectrograph slit by means of an achromatic condensing lens, whose aperture was such that the collimating lens of the spectrograph was completely filled with light. On each photographic plate exposed there were up to 15 caesium spectra, as well as spectra of a standard lamp and of density marks; the exposure time was 10 minutes in each case. The spectrum of the standard lamp, whose colour temperature was known, was photographed under optical conditions identical with those used in photographing the caesium spectra. In the course of the experiments two standard lamps were used, providing a check on the calibration of each. Five density marks were obtained on each plate by exposing the spectrograph slit to the radiation reflected from a diffusing surface coated with magnesium oxide for different widths of the slit.

Photometry was carried out manually on a Zeiss microphotometer, over the wavelength region 4985–4170 Å of each spectrum. The projected height of the photometer slit, which was set parallel to the image of the spectrograph slit, was such that the radiation taken into account in the photometry of the caesium spectra was emitted by a central 'slice' of the discharge tube of thickness 0·4 mm. Characteristic curves were constructed from the density marks, for each wavelength investigated. Intensities at different wavelengths were related by means of the intensity distribution in the standard lamp spectrum, which was taken to be that of a black body at the colour temperature. In fig. 2 is shown the intensity distribution in the 6P recombination continuum of caesium, for particular discharge conditions; for convenience in calculating T_1 from eqn. (3), $\log\{\nu J(\nu)\}$ is plotted as a function of ν.

In curve (*a*) of fig. 2 there is a distinct non-linearity in the relationship between $\log\{\nu J(\nu)\}$ and frequency ν. The relationship became more nearly linear for larger values of the caesium vapour pressure; this fact, coupled with the shape of the curve in relation to the frequency of an intense principal series doublet (4593, 4555 Å) of the caesium arc spectrum, suggested that the non-linearity was due to light scattered from this doublet. This interpretation was confirmed by experiments in which the radiation from the discharge was passed through a Pyrex glass absorption cell containing caesium vapour at a pressure of 10^{-2} mm Hg, and situated between the discharge tube and the condensing lens; the vapour in this cell absorbed the radiation of the principal series doublet. Curve (*b*), fig. 2 shows the relationship between $\log\{\nu J(\nu)\}$ and ν for the same discharge conditions as for curve (*a*), but with the absorption cell in position. It follows that the

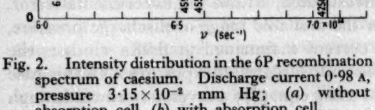


Fig. 2. Intensity distribution in the 6P recombination spectrum of caesium. Discharge current 0·98 A, pressure $3{\cdot}15\times10^{-2}$ mm Hg; (*a*) without absorption cell, (*b*) with absorption cell.

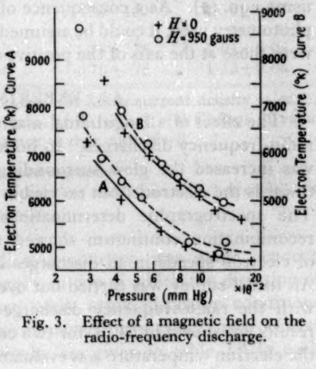


Fig. 3. Effect of a magnetic field on the radio-frequency discharge.

non-linearity in (*a*) is caused by light scattered from the intense blue doublet, and that there is a Maxwellian distribution of electron velocities in the discharge, within the experimental error. The absorption cell was discarded in subsequent experiments, as it was found that the cell windows became rapidly discoloured under the action of the caesium vapour; the discoloration of the glass wall of the positive column was much less rapid. In the determination of the value of T_1 with no absorption cell, values of $J(\nu)$ in the neighbourhood of the blue doublet were discarded.

The slight progressive discoloration of the glass wall of the positive column gave rise to errors in the spectrographic determinations of T_1; the absorption in the tube wall was such that falsely low values of T_1 were obtained. After an initial set of experiments, changes produced in the value of T_1 by a variation in the discharge conditions were determined more accurately using eqn. (5). If the parameters of the discharge in the two cases to be compared are denoted by subscripts a and b, we have only to determine the ratio $J_a(\nu)/J_b(\nu)$ at a number of values of the frequency in order to obtain $T_{1a}^{-1}-T_{1b}^{-1}$; this procedure does not involve the spectrum of the standard lamp, and is also independent of any discoloration in the walls of the positive column, provided that this may be assumed to have remained constant during the exposure of the two spectra. The absolute value of one of T_{1a}, T_{1b} was taken as the mean of a number of determinations using eqn. (3).

Relative values of the electron density in two discharges were determined using eqn. (4). As a consequence of the optical arrangement and the method of photometry used, it could be assumed that the values of electron density concerned were those at the axis of the positive column.

§ 3. EXPERIMENTAL RESULTS

The effect of a longitudinal magnetic field was investigated for both d.c. and radio-frequency discharges. In both cases, as the intensity of the magnetic field was increased the glow surrounding each of the electrodes was 'compressed' towards the electrode, but no visible effect was produced in the positive column. The spectrographic determinations of the distribution of intensity in the recombination continuum showed that there was a Maxwellian distribution of electron speeds in all discharges investigated, within the experimental error. An initial survey was carried out over the available range of discharge pressure, with the radio-frequency discharge current maintained at 0·98 A r.m.s.; the results are shown in fig. 3, for two consecutive plates of exposures. In this case the electron temperature was evaluated using eqn. (3); curves are drawn through the values of T_1 for zero field, and for a longitudinal magnetic field of intensity 950 ± 40 gauss, the value used throughout the survey. An additional experiment with radio-frequency excitation was carried out on tube No. 3, with a mean current density of 5 A cm^{-2} at a frequency of 6·65 Mc/s. At a pressure of 0·078 mm Hg the value of T_1 was increased 175 ± 100° K by a magnetic field of intensity $H=1450$ gauss, from its initial value of 4000° K for $H=0$. This increase was determined with the use of eqn. (5). In all the experiments in which the discharge was excited by radio-frequency energy, the application of a longitudinal magnetic field produced no measurable change in the axial value of the electron density N_1. It is estimated that a change in the value of N_1 of 5% or more would have been detected.

The remaining experimental results for the effect of a longitudinal magnetic field all concern discharges carrying a direct current, in tubes Nos. 2 and 3; the results are summarized in table 1.

Table 1. Characteristics of the d.c. Discharge

(1)	(2)	(3)	(4)	(5)	(6)	(7)
0·0285	4*	5·0	5450	1100	+450±200	0·82±0·10
	4*	5·0	5450	1500	+575±200	0·77±0·10
0·0485	1	5·0	4600	650	+450±150	0·76±0·04
	1	5·0	4600	1150	+800±150	0·62±0·04
	1	5·0	4600	1450	+1050±150	0·57±0·03
	1	5·0†	4700	600	−100±150	1·28±0·07
	1	5·0†	4700	1150	−300±150	1·74±0·09
	1	5·0†	4700	1450	−300±150	1·88±0·10
0·0785	4	2·5	3750	1400	+100± 75	0·99±0·05
	1	5·0	4200	650	+150±100	0·86±0·05
	2	5·0	4200	1050	+300±100	0·87±0·05
	2	5·0	4200	1400	+400±100	0·84±0·05
	4*	5·0	4650	1450	+200±150	0·85±0·10
	1*	5·0	4650	1450	+350±200	0·84±0·10
	4	5·0†	4200	400	+75±100	1·00±0·05
	1	5·0†	4200	650	0±100	1·07±0·06
	2	5·0†	4200	1450	0±100	1·22±0·07
	1	5·0†	4200	1450	−50±100	1·30±0·07
0·11	1	5·0	3625	1450	+75± 75	0·91±0·05
	1	5·0†	3625	1450	+75± 75	1·13±0·06

(1) Pressure (mm Hg); (2) No. of determinations; (3) Mean current density (A/cm^2); (4) Mean value of T_1 for $H=0$ (° K); (5) Mean value of H (gauss); (6) Change in value of T_1 (° K); (7) Axial electron concentration, relative to that with $H=0$.

* Caesium vapour absorption cell used in the spectrography.

† Electrode nearer appendix tube (cf. fig. 1) operated as anode; in all other cases, this electrode was operated as the cathode of the discharge.

With zero longitudinal magnetic field, a comparison was made between values of T_1 and of N_1 for two d.c. discharges, in one of which the direction of flow of the discharge current was reversed while other discharge conditions remained identical. Similar data were also obtained for a comparison of d.c. and radio-frequency discharges. These results are given in table 2: in each case the discharge current was 0·98 A (r.m.s. in the case of radio-frequency excitation), corresponding to a mean current density of 5 A/cm^2. Changes in the value of T_1 which are given in tables 1 and 2 were determined by the method of eqn. (5).

Estimates of the probable errors in the experimental values of N_{10} and of changes in T_1 are included in tables 1 and 2. As the vapour pressure increased the estimated error became smaller in the case of changes in the value of T_1, because of the reduced effect of light scattered from the blue doublet on the intensity distribution in the recombination continuum. The presence of the absorption cell gave rise to greater errors, due to the progressive discoloration of the windows of the cell.

Measurements were also made of the change produced in the value of the total potential drop across the discharge tube, when it was subjected to a longitudinal magnetic field. In all cases the change in potential difference across the tube was less than 3% for a value of the field $H = 1500$ gauss. In general the potential difference was increased by a magnetic field of this value, but the increase was not always a monotonic function of H.

Table 2. Comparison of d.c. and high-frequency Discharges in the Absence of Magnetic Fields

Pressure (mm Hg)	Frequency	Derived value of T_1 (° K)	Rel. value of axial electron conc.
0·0485	d.c.	4600*	1·00
	d.c.†	4700 ± 150	$0{\cdot}81 \pm 0{\cdot}04$
	4·00 Mc/s	4625 ± 150	$0{\cdot}96 \pm 0{\cdot}05$
	6·65 Mc/s	4525 ± 150	$0{\cdot}97 \pm 0{\cdot}05$
0·0785	d.c.	4200*	1·00
	d.c.†	4200 ± 100	$0{\cdot}90 \pm 0{\cdot}05$
	6·65 Mc/s	4000 ± 100	$1{\cdot}07 \pm 0{\cdot}06$
0·11	d.c.	3625*	1·00
	d.c.†	3625 ± 75	$0{\cdot}95 \pm 0{\cdot}05$
	6·65 Mc/s	3500 ± 75	$1{\cdot}08 \pm 0{\cdot}06$

* Reference value, derived by means of eqn. (3).

† Electrode nearer appendix tube (cf. fig. 1) operated as anode; in all other d.c. discharges above, this electrode was operated as the anode of the discharge.

§ 4. DISCUSSION

From the spectroscopic investigations of the recombination radiation the effect of a longitudinal magnetic field on values of T_1 and N_1 was found to depend upon the method of excitation of the caesium discharge: it differed for the two cases of radio-frequency and d.c. discharges, and in the latter case depended upon the direction of flow of the direct current relative to the appendix tube containing the liquid caesium. When the electrode nearer the appendix tube was operated as the cathode of a d.c. discharge, the value of T_1 increased and that of N_{10} decreased with increasing intensity of the magnetic field; when this electrode was operated as anode, the reverse was the case. The effect was most marked at a pressure of 0·0485 mm Hg, as can be seen from the results of table 1. The effect of the magnetic field was less pronounced at a lower value of the discharge current.

From investigations of d.c. discharges with no magnetic field present (table 2), the values of T_1 and N_1 in the positive column were found to depend upon the direction of flow of the discharge current in this case also. Experiments on rare gases (Rüttenauer 1922) have shown that the flow of a discharge current sets up a pressure gradient along the positive column; these results have been interpreted by the theory of Langmuir (1923). Since it was found here that, for a given value of the caesium discharge current, the value of T_1 decreased and that of N_1 increased with increasing pressure, we may attribute the results of table 2 for d.c. discharges to the production of a similar pressure gradient in the caesium discharge. The calculated values of the gradient were of the same order as those observed by Rüttenauer, with the pressure greater at the anode in each case. It should be remarked that a pressure gradient could be set up in the caesium vapour, because the main body of the discharge tube was kept at a temperature greater than that of the appendix tube.

The spectroscopic results for d.c. discharges in a magnetic field (table 1) are then consistent with this hypothesis: the longitudinal magnetic field, in conjunction with the flow of discharge current, produces a pressure gradient along the positive column directed so that the pressure is greater at the cathode, i.e. in opposition to the gradient existing with no magnetic field present. It would be of interest to carry out physical measurements of the pressure in search of such a gradient, and also to attempt an interpretation in terms of Langmuir's theory.

When radio-frequency discharges were subjected to a longitudinal magnetic field the value of N_{10} was unchanged over the whole range of discharge pressure while the value of T_1 was increased (fig. 3). A determination by the more accurate method of eqn. (5) showed the value of T_1 to be increased by approximately 5% by a field of 1450 gauss, at a pressure of 0·078 mm Hg.

According to the theory of Tonks (1939) the electron temperature T_1 is decreased by the presence of a longitudinal magnetic field. This decrease is ascribed to the decreased rate of diffusion of the electrons towards the wall of the positive column. If we use the expression due to Townsend (1912) for the coefficient of diffusion of electrons normal to a magnetic field of intensity H:

$$D(H) = D(0)/(1+\omega^2\tau^2), \qquad \ldots\ldots(6)$$

where $D(0)$ is the diffusion coefficient in the absence of a magnetic field, ω is the gyrofrequency and τ the mean free time of flight of the electrons between collisions, then approximate calculations show that the value of T_1 should theoretically be decreased by 10% by a magnetic field of intensity only 200 gauss, at a pressure of 0·049 mm Hg. In these calculations, use was made of experimental data for similar discharges in caesium obtained by Mohler (1938 b). Comparing with this the results described here for radio-frequency discharges, there is wide disagreement. The latter results were chosen for comparison in preference to those for d.c. discharges, because of the possibility of effects due to pressure gradients in d.c. discharges in a longitudinal magnetic field.

Townsend (1938) has shown that the expression (6) is confirmed experimentally for electrons diffusing through a neutral gas, for which case it was originally derived. There thus appears to be a fundamental difference from the Townsend case in the diffusive motion of electrons normal to a magnetic field, when they are accompanied by an equal density of positive ions which have equal drift velocity normal to the magnetic field.

In conclusion, the results of table 2 for zero intensity of the magnetic field enable a comparison to be made between d.c. discharges carrying a given current, and radio-frequency discharges at the same pressure and carrying the same r.m.s. current. Neglecting effects due to the presence of a slight pressure gradient in the positive column of the d.c. discharges, no significant differences were found in the values of T_1 and N_1 in the two cases. Comparisons between discharges of these two types have been made previously by Beck (1935), who used a modified Langmuir probe technique to measure the electron temperature in both d.c. and radio-frequency discharges (frequency 100 Mc/s). In the positive column of a discharge in mercury vapour, Beck found the value of T_1 to be substantially the same for d.c. and radio-frequency discharges of low current density over a wide range of pressure. The results reported here are in agreement with these findings, and thus lend support to Beck's method of determining electron temperatures in radio-frequency discharges.

ACKNOWLEDGMENTS

I should like to express my thanks to Dr. H. G. Kuhn, under whose supervision these experiments were carried out in the Clarendon Laboratory, and to Dr. P. C. Thonemann, who suggested an investigation of discharges in magnetic fields using the spectroscopic method described above; both gave advice and assistance throughout. I am also indebted to Dr. A. von Engel for valuable discussions. This work was carried out during the tenure of a Rhodes Scholarship.

REFERENCES

BECK, H., 1935, *Z. Phys.*, **97**, 355.
BOECKNER, C., 1931, *Bur. Stand. J. Res.*, *Wash.*, **6**, 277.
DITCHBURN, R. W., and GILMOUR, J. C., 1941, *Rev. Mod. Phys.*, **13**, 310.
FINKELNBURG, W., 1938, *Kontinuierliche Spektren* (Berlin : Springer); 1948, *Physics of the Electron Shells* (Wiesbaden : Office of Military Government for Germany, F.I.A.T.), p. 18.
GIOVANELLI, R. G., 1948, *Aust. J. Sci. Res.* A, **1**, 275.
JANCKE, H. O., 1936, *Z. Phys.*, **99**, 169.
LANGMUIR, I., 1923, *J. Franklin Inst.*, **196**, 751.
MOHLER, F. L., 1936, *Bur. Stand. J. Res.*, *Wash.*, **17**, 849; 1937, *Ibid.*, **19**, 447; 1938 a, *Ibid.*, **21**, 697; 1938 b, *Ibid.*, **21**, 873.
MOHLER, F. L., and BOECKNER, C., 1929, *Bur. Stand. J. Res.*, *Wash.*, **2**, 489.
RÜTTENAUER, A., 1922, *Z. Phys.*, **10**, 269.
TONKS, L., 1939, *Phys. Rev.*, **56**, 360.
TOWNSEND, J. S., 1912, *Proc. Roy. Soc.* A, **86**, 571; 1938, *Phil. Mag.*, **25**, 459

Phase Stability of the Microtron

BY C. HENDERSON, F. F. HEYMANN AND R. E. JENNINGS
University College, London

Communicated by H. S. W. Massey; MS. received 3rd September 1952, *and read before the Society at University College, London,* 1*st February* 1952

ABSTRACT. The limits of phase and energy within which electrons can be stably accelerated in a microtron are calculated for a number of voltages by two different methods. The effect of small changes in the magnetic field are considered and the energy and phase of electrons after the first transit are given for different resonator gaps and voltages. Finally, the design of resonator and the operating conditions likely to give maximum output from this type of machine are discussed.

---

§ 1. INTRODUCTION

THE conventional cyclotron fails to accelerate electrons to high energies because the relativistic increase of mass causes a departure from synchronism between the orbit period and the period of the radio-frequency supply. Veksler (1945) pointed out that when the peak voltage across the accelerating gap has certain discrete values, corresponding generally to integral multiples of the electron rest mass, then a resonant acceleration which is in some ways analogous to that occurring in the cyclotron can take place without the complication of frequency modulation or varying magnetic fields.

The microtron, or electron cyclotron, is the practical form of such an accelerator, and the first machine was built in Ottawa (Redhead *et al.* 1950). A similar machine has recently been completed here. It consists essentially of a resonant cavity situated in a uniform magnetic field. The cavity is energized by a pulsed magnetron on a wavelength of about 10 cm, and in this way it is possible to obtain a very high electric field across the lips of the cavity. The electrons, obtained by field emission from the lips of the cavity and accelerated each time they pass through the cavity, describe orbits of increasing radius which all have a common tangent through the resonator, as shown in fig. 1. Apart from the first orbit, the condition necessary for resonant acceleration is that the increase in energy obtained by an electron each time it crosses the gap of the resonator shall be such as to make the electron take a whole number of periods longer for each successive orbit. Many different modes are possible, but this paper is only concerned with the mode in which the electron gains energy corresponding to one rest mass per transit through the resonator, which has a peak voltage slightly above 0·5 MV across its lips, and the magnetic field is so adjusted that the time for each successive orbit increases by one period of the radio-frequency field. Redhead, LeCaine and Henderson (1950) discuss the range of phase angles for which it is possible to obtain stable acceleration, but it is thought that the criterion used in their calculation is incorrect. This is borne out by the fact that stable acceleration was found in practice to be possible over much wider limits than those they predicted.

The question of phase stability is of importance, particularly as the source of electrons is the field emission from the lips of the resonant cavity. This source although very convenient, has the disadvantage that many of the electrons are emitted with the wrong phase and/or energy for stable acceleration, and Redhead *et al.* found in their machine that only about 1% of the current in the first orbit reached the second and later orbits. As the maximum number of electrons is emitted when the voltage across the gap is near its peak, it is important to design the resonant cavity so that these electrons will have suitable energy and phase for stable acceleration.

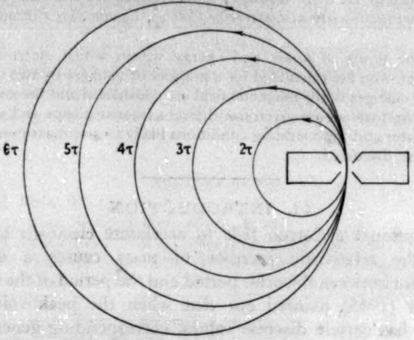


Fig. 1. Resonator and electron orbits.

The range of phase angles for stable acceleration has also been considered by Itoh and Kobayashi (1950), and the results they obtain are in substantial agreement with those obtained here. The permissible variation in energy for stable acceleration was not discussed, however, nor the way in which best use could be made of electrons obtained by field emission. In their paper they suggested injecting high-energy electrons into the resonant cavity, but in practice this would be difficult to do in this type of machine.

§ 2. LIMITS OF PHASE STABILITY

In the following calculations it is assumed that the electric field between the lips of the resonator is everywhere parallel to the common tangent to the orbits and varies sinusoidally with time. It can then be shown that for transits other than the first the energy ϵ gained by an electron in crossing the gap is given by

$$\epsilon = \boldsymbol{e}V_0\left[\frac{\sin(\alpha_2-\alpha_1)/2}{(\alpha_2-\alpha_1)/2}\right]\cos(\alpha_1+\alpha_2)/2 \qquad \ldots\ldots(1)$$

where the voltage across the gap is represented by $V=V_0\cos\alpha$, $\boldsymbol{e}$ is the electronic charge and α_1, α_2 are the phase angles at which particle enters and leaves the gap.

The expression in the square brackets varies very slowly for values of $\alpha_2-\alpha_1$ between 30° and 40° and can be taken as constant. (All calculations are for $\alpha_2-\alpha_1=37°$ except in §4). It is convenient to use the mean phase angle in the calculations and to write (1) in the form

$$\epsilon = \boldsymbol{e}V'\cos\bar{\alpha} \qquad \ldots\ldots(2)$$

where $V'=V_0\left[\frac{\sin(\alpha_2-\alpha_1)/2}{(\alpha_2-\alpha_1)/2}\right]$ and $\bar{\alpha}=(\alpha_1+\alpha_2)/2$.

In the particular mode used, at each transit the electron should cross the gap of the cavity after the peak and at a mean phase angle $\bar{\alpha}$ such that its increase in energy ϵ is equal to eV_{R}, the rest energy of the electron. If the magnetic field is then in correct adjustment, these increments in energy make the time for each successive orbit increase by one period of the radio-frequency field. An electron which behaves exactly in this way will be referred to as the 'resonant' electron. The object of the following calculations is to determine how far an electron can deviate from these ideal conditions without falling out of synchronism, and it is convenient to make the calculations relative to the 'resonant' electron.

Suppose that an electron leaves the resonator after the *n*th transit at a phase angle $\bar{\alpha}_n$ with an energy error of $\Delta\epsilon_n$ relative to the resonant electron. As an increase in energy of eV_{R} makes an electron take an extra period of the radio-frequency field to complete an orbit and, as the time to complete an orbit is directly proportional to the total energy of the electron, the electron considered will cross the gap at a phase angle $\bar{\alpha}_{n+1}$ given by

$$\bar{\alpha}_{n+1} = \bar{\alpha}_n + 2\pi\Delta\epsilon_n/eV_{\mathrm{R}}. \qquad \ldots\ldots(3)$$

The energy error as it leaves the resonator after the $(n+1)$th transit will be given by

$$\Delta\epsilon_{n+1} = \Delta\epsilon_n + eV_{\mathrm{R}}\left[\frac{V'}{V_{\mathrm{R}}}\cos\bar{\alpha}_{n+1} - 1\right]. \qquad \ldots\ldots(4)$$

Equations (3) and (4) may be combined to give the equation

$$\bar{\alpha}_{n+2} - 2\bar{\alpha}_{n+1} + \bar{\alpha}_n = 2\pi\left[\frac{V'}{V_{\mathrm{R}}}\cos\bar{\alpha}_{n+1} - 1\right]. \qquad \ldots\ldots(5)$$

This equation is similar to the equation expressing the phase oscillations of the synchrotron. In the case of the microtron, however, the transition to the corresponding differential equation is not allowable, since changes in $\bar{\alpha}_n$ from orbit to orbit are not in general small enough to warrant replacing the left-hand side by its corresponding differential coefficient. Consequently the equation was solved numerically.

The interesting result of such calculations was the fact that, provided the initial values of $\Delta\epsilon_n$ and $\bar{\alpha}_n$ lay within certain limits, the points representing values of $\Delta\epsilon_n$ and $\bar{\alpha}_n$ after subsequent transits were found to lie on a closed curve, as shown in fig. 2 (outer curve) for example.

This curve is the limiting case for $V_0 = 570$ kv, for if any point outside this curve is chosen for the initial values of $\Delta\epsilon_n$ and $\bar{\alpha}_n$, it is found that points representing subsequent transits do not lie on a closed curve and that the electron soon becomes unstable. If, on the other hand, any initial point is chosen inside the limiting curve, points representing subsequent transits are found to lie on a smaller version of the limiting curve as shown. The region inside the limiting curve therefore represents phase stable conditions.

A similar set of curves for $V_0 = 530$ kv are shown in fig. 3, including the transits of an electron which eventually becomes unstable. At transit No. 9 the phase angle is such that the electron crosses the resonator when $V'\cos(-\bar{\alpha})$ is approximately equal to V_{R}. Since the electron still has a negative energy error after this transit, the next time it crosses the gap of the resonator earlier in the cycle, when $V'\cos(-\bar{\alpha}) < V_{\mathrm{R}}$, so that its energy error becomes increasingly negative and it drops out of synchronism.

In fig. 4 are plotted the limiting curves for a number of values of V_0. It is shown in the Appendix that resonant acceleration is only possible for values of V_0 between 520 and 614 kv, and it can be seen that the curves decrease in area as they approach these limits.

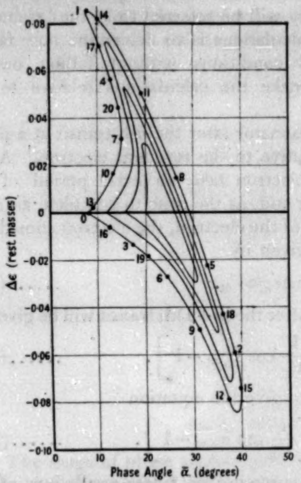


Fig. 2. Phase stable regions for $V_0 = 570$ kv. Transit numbers are given on the outermost curve only.

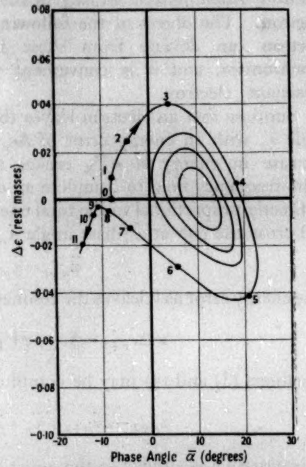


Fig. 3. Phase stable regions for $V_0 = 530$ kv. The outer curve shows the successive transits of an unstable electron.

When drawing these curves it is difficult to lay down an exact criterion for instability. In many cases it is found that, starting with a point just outside a limiting curve, the subsequent points, whilst not quite lying on a closed curve, still appear to represent stable conditions. Also, at 550 kv, although it is possible to find a succession of points which show no sign of becoming unstable, they do not lie on a closed curve at all. The curves in fig. 4 probably give a conservative estimate of the phase stable regions, because the limiting curve was taken as the one for which a smooth closed curve could be drawn through *all* the points.

The way in which the electron corresponding to the outer curve of fig. 3 becomes unstable suggests a different approach in which an exact criterion of instability can be given. In fig. 5 $(V' \cos \bar{\alpha} - V_R)$ is shown plotted against $\bar{\alpha}$, and it can be seen that if electrons cross the gap of the resonator at a phase $-\bar{\alpha}_A$ (where $V' \cos(-\bar{\alpha}_A) = V_R$) with negative energy errors as they leave the gap, then they will rapidly drop out of synchronism. Thus electrons represented by points along the line AB are certainly unstable. These electrons are represented for the previous transit by the line AC. This line, which is thus known to represent unstable electrons, is used as the starting point in this second method of calculation. It is convenient to adjust the scales so that the line AC is at 45°, $\bar{\alpha}$ again being the mean transit angle and $\Delta\epsilon$ the energy error as the electron leaves the resonant cavity.

Starting from this line, values of $\Delta\epsilon$ and $\bar{\alpha}$ for the previous transit can quickly be found graphically and are shown as curve 1 in fig. 6. Any electrons represented by points in the region to the right of this curve would, after their next transit, be represented by points below the starting line, showing that this region corresponds to unstable conditions. On successively drawing further curves corresponding to earlier transits the unstable region increases quite quickly to a limit, the area remaining inside corresponding to stable conditions. If the curves are examined closely it will be seen that thin streaks representing stable conditions are found in the unstable regions, but these are not important to the general picture. In figs. 6 to 8 three examples are given. For 530 and 570 kv the two methods are seen to be in quite good agreement, the region obtained by the first method being indicated by the dashed line. The second method could probably be brought into better agreement with the first if a larger number of curves were drawn, but only the important ones for a comparatively small number of transits have been included for the sake of clarity.

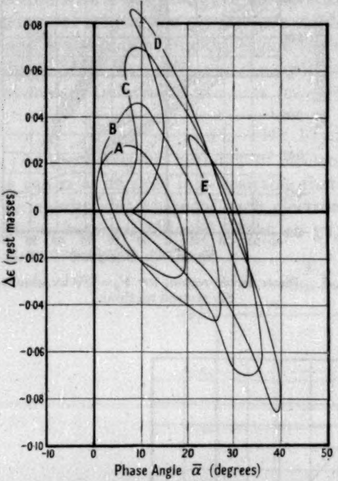


Fig. 4. Phase stable regions. A 530 kv, B 540 kv, C 560 kv, D 570 kv, E 590 kv.

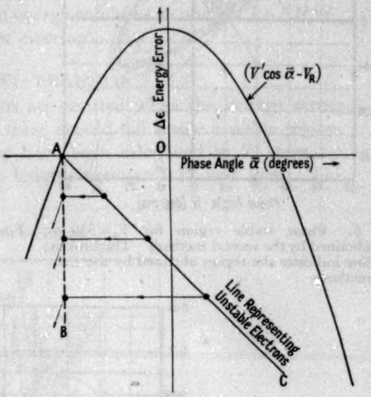


Fig. 5.

For 550 kv the stable region obtained by the second method has an extremely awkward shape and quite large regions appear to be isolated by streaks of instability. This probably accounts for the fact that a closed curve for 550 kv was not obtained by the first method. The succession of apparently stable points obtained by the first method, however, do lie in the stable regions of fig. 7.

By either of the two methods employed a fairly accurate estimate of the stable region is thus obtained. The results are in general agreement with the limits of phase for stable acceleration given by Itoh and Kobayashi, who, unfortunately, do not include details of their method of calculation.

It must be remembered that in machines of this type where the number of transits are small, many electrons outside the stable regions will still be fully accelerated as they will not make sufficient transits to reach the point where they would drop out of synchronism.

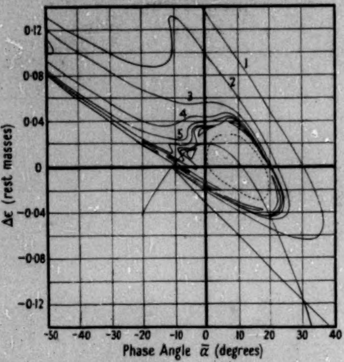


Fig. 6. Phase stable region for $V_0 = 530$ kv obtained by the second method. The broken line indicates the region obtained by the first method.

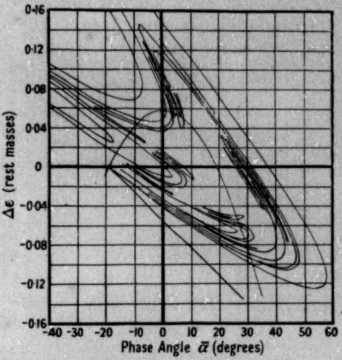


Fig. 7. Phase stable region for $V_0 = 550$ kv obtained by the second method.

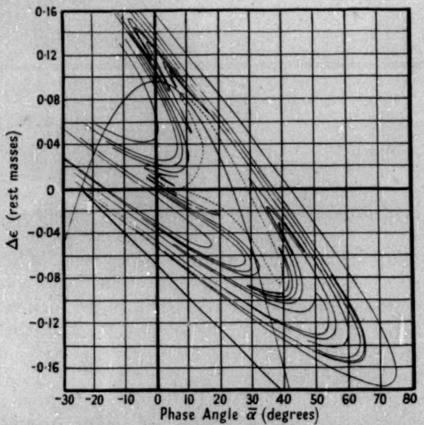


Fig. 8. Phase stable region for $V_0 = 570$ kv obtained by the second method. The broken line indicates the region obtained by the first method.

§ 3. EFFECTS OF VARIATIONS OF THE MAGNETIC FIELD

Before considering the first transit, the effect of small variations in the magnetic field will be considered. As the magnetic field cannot change appreciably during the short time taken to accelerate an electron, it is sufficient to calculate the effect of a small change in field from one steady state to another.

The time for an electron to complete an orbit is given by $T=2\pi U/Hec^2$ where U is the total energy of electron, an integral multiple of $\boldsymbol{e}V_{\mathrm{R}}$ for the mode considered above and H is the strength of magnetic field.

The value of H for the resonant electron, which takes a whole number of periods to complete an orbit, is denoted by H_{R}. If H changes from H_{R} to $H_{\mathrm{R}}+\Delta H$, for an electron to cross the gap at the correct phase, its total energy must become an integral multiple of $\boldsymbol{e}(V_{\mathrm{R}}+\Delta V)$ such that $V_{\mathrm{R}}/H_{\mathrm{R}}=(V_{\mathrm{R}}+\Delta V)/(H_{\mathrm{R}}+\Delta H)$. If the previous calculations were now repeated, relative to a 'new' resonant electron with a rest energy of $\boldsymbol{e}V_{\mathrm{R}}(1+\Delta H/H_{\mathrm{R}})$, and which gains that amount of energy per transit, the set of phase stable regions obtained would have the same shape as previously, but the region for a voltage V would be that labelled for a voltage $VH_{\mathrm{R}}/(H_{\mathrm{R}}+\Delta H)$, and the errors in energy would be in units of $1+\Delta H/H_{\mathrm{R}}$ rest masses (relative to the new resonant electron).

§ 4. THE FIRST TRANSIT

As the maximum number of electrons are emitted when the voltage across the gap is at its peak it is desirable that these should fall inside a stable region. The energy and phase of such electrons have been calculated in 5° steps for voltages between 530 and 590 kv on gap lengths between 0·75 and 1 cm. The results are shown in fig. 9.

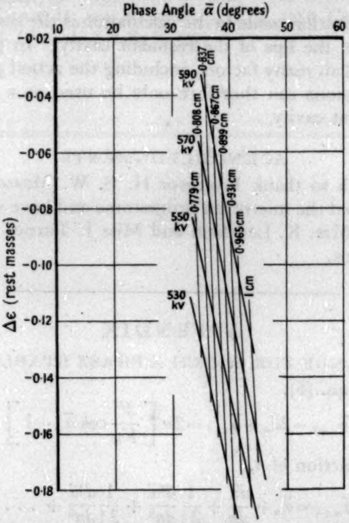


Fig. 9. Phase $\bar{\alpha}$ and energy error of electrons, emitted at the peak cavity voltage, after their first transit.

Here the phase has been expressed relative to the mid-point of later transits, so that the points can be compared directly with the boundary curves. On doing so, it is seen that in general the points lie outside the corresponding stable regions. The situation can be improved however by making small changes in the magnetic field. If the field is changed from H_R to $H_R+\Delta H$, an electron after the first transit should have a total energy of $2eV_R(1+\Delta H/H_R)$, so that a previous energy deficiency of $\Delta\epsilon$ rest masses now becomes $\Delta\epsilon+2\Delta H/H_R$ rest masses, or $(\Delta\epsilon+2\Delta H/H_R)[H_R/(H_R+\Delta H)]$ in units of $1+\Delta H/H_R$ rest masses. It is the expression in round brackets which is of importance, and the chief effect of small variations of the magnetic field is to move the first transit points parallel ot the energy error axis.

One operating point which has good possibilities is found for a field 2% below normal when electrons accelerated by a peak voltage of 560 kv on a 0·87 cm gap lie well inside the phase stable region for 560/(1 − 0·02) kv, i.e. 570 kv. Also many electrons emitted before the peak, with smaller values of $-\Delta\epsilon$ and $\bar{\alpha}$, will be stably accelerated under these conditions, so that the fraction of electrons going on from the first orbit should be high. The shunt resistance of such a small gap will, however, be rather low, requiring a more powerful radio-frequency source than would otherwise be necessary.

§ 5. CONCLUSION

The conditions for which a large beam current should be obtained from a microtron have been discussed, and it has been shown that small variations in the strength of the magnetic field about its resonant value can assist in achieving these conditions.

The chief simplification made in the calculation is the assumption of a parallel electric field between the lips of the resonant cavity. In practice the shape of this field will depend on many factors, including the actual profile chosen for the lips. The stable regions can therefore only be used as a general guide in the design of the resonant cavity.

ACKNOWLEDGMENTS

The authors wish to thank Professor H. S. W. Massey and Dr. E. H. S. Burhop who instigated the microtron programme and have shown a keen interest in its progress, and Mrs. K. Ledsham and Miss J. Turner who helped with the numerical calculations.

APPENDIX

THE VOLTAGE RANGE FOR WHICH A PHASE STABLE REGION EXISTS

We have, from eqn. (5),

$$\bar{\alpha}_{n+1}-2\bar{\alpha}_n+\bar{\alpha}_{n-1}=2\pi\left[\frac{V'}{V_R}\cos\bar{\alpha}_n-1\right].$$

Expanding $\bar{\alpha}$ as a function of n,

$$\bar{\alpha}_{n+1}=\bar{\alpha}_n+\frac{d\bar{\alpha}}{dn}+\frac{1}{2!}\frac{d^2\bar{\alpha}}{dn^2}+\frac{1}{3!}\frac{d^3\bar{\alpha}}{dn^3}+\dots.$$

and

$$\bar{\alpha}_{n-1}=\bar{\alpha}_n-\frac{d\bar{\alpha}}{dn}+\frac{1}{2!}\frac{d^2\bar{\alpha}}{dn^2}-\frac{1}{3!}\frac{d^3\bar{\alpha}}{dn^3}+\dots.$$

so that

$$\bar{\alpha}_{n+1}-2\bar{\alpha}_n+\bar{\alpha}_{n-1}=2\left[\frac{1}{2!}\frac{d^2\bar{\alpha}}{dn^2}+\frac{1}{4!}\frac{d^4\bar{\alpha}}{dn^4}+\dots\right].$$

Now consider oscillations about $\bar{\alpha}_0$, where $\bar{\alpha}_0$ is given by $\cos\bar{\alpha}_0=V_R/V'$.

If $\bar{\alpha}-\bar{\alpha}_0$ is denoted by θ, eqn. (5) can be written

$$2\left[\frac{1}{2!}\frac{d^2\theta}{dn^2}+\frac{1}{4!}\frac{d^4\theta}{dn^4}+\ldots\right]=2\pi\left[\frac{V'}{V_R}\cos\theta\cos\bar{\alpha}_0-\frac{V'}{V_R}\sin\theta\sin\bar{\alpha}_0-1\right]$$

$$=2\pi[\cos\theta-\tan\bar{\alpha}_0\sin\theta-1].$$

For small oscillations for which $\sin\theta$ may be replaced by θ, and $\cos\theta\simeq 1$, we have

$$2\left[\frac{1}{2!}\frac{d^2\theta}{dn^2}+\frac{1}{4!}\frac{d^4\theta}{dn^4}+\ldots\right]=-2\pi\theta\tan\bar{\alpha}_0.$$

To solve this equation put $\theta=A\sin Kn$, giving

$$2\left[-\frac{1}{2!}K^2+\frac{1}{4!}K^4+\ldots\right]=-2\pi\tan\bar{\alpha}_0,$$

i.e. $-2+2\cos K=-2\pi\tan\bar{\alpha}_0$. Hence the limits within which small oscillations are possible are given by $2\pi\tan\bar{\alpha}_0=0$ and 4.

Therefore $\bar{\alpha}_0$ lies between 0 and 32·5° and V' between 511 to 606 kv. When the transit angle is in the neighbourhood of 37°, V_0 the actual voltage across the cavity lies between 520 and 614 kv peak.

REFERENCES

ITOH, J., and KOBAYASHI, D., 1950, *Col. Fac. Sci., Osaka University*, B, 11.

REDHEAD, P. A., LECAINE, H., and HENDERSON, W. J., 1950, *Canad. J. Res.* A, **28**, 73.

VEKSLER, V., 1945, *J. Phys., U.S.S.R.*, **9**, 153.

Carrier Injection and Extraction in Lead Sulphide

By P. C. BANBURY
Physics Department, University of Reading

Communicated by R. W. Ditchburn; MS. received 24th March, and in final form 3rd September 1952

ABSTRACT. A method of measuring changes of minority carrier concentration in a semiconductor is discussed. The method is applied to the investigation of injection in specimens of p-type lead sulphide which show transistor action. There is evidence of electron injection at both ' point ' metal contacts and contacts of area up to 10^{-4} cm^2. In the latter case an additional effect is observed which is interpreted as the reverse of injection, i.e. electron extraction.

---

§ 1. METHOD OF INVESTIGATION OF CARRIER INJECTION

The occurrence of carrier injection in transistor materials was originally inferred from measurements of probe potentials (Bardeen and Brattain 1948). The interpretation of results obtained by such methods may be complicated by changes of 'floating potential', in the sense used by Bardeen (1950), which accompany changes in minority carrier concentration. The following alternative method of investigation is believed to overcome this difficulty and give relative values of injected carrier density.

A schematic diagram of electron energies for a contact in equilibrium between a metal and an n-type semiconductor is shown, for the purpose of defining symbols, in fig. 1. For a given contact, in the absence of nearby injection, the application of a fixed reverse voltage in the saturation range will give rise to a certain 'collector' current I_c. Two possible contributions to this current must be considered, corresponding to a flow of electrons or holes in the semiconductor. The current of electrons flowing into the conduction band at a fixed voltage and temperature will depend only on W_1 (fig. 1) and the contact area. On the other hand, the hole current flowing into the metal is expected to be independent of W_1, but limited by the contact area and by the hole replacement mechanism in the surrounding semiconductor. Differences in the value of I_c when the contact is moved from place to place may be attributed to changes of electron current due to variations in W_1, or of hole current due to variations in the replacement process, assuming the contact areas to be equal.

In the presence of a nearby emitting electrode a certain number of the injected holes, as determined by lifetime, diffusion and drift conditions, will reach the contact under discussion. A change in open circuit floating potential will ensue, corresponding to a change in W_2, but W_1 is not expected to change, and will be taken as constant. Hence the electron current for reverse bias will be insensitive to the presence of these added holes. There will, however, be a change in hole current ΔI_c directly proportional to the added hole concentration near the contact. ΔI_c should be independent of I_c, and a direct measure of the injected carrier concentration, provided that electric fields in the semiconductor are small. It should therefore be possible to move the probe from place to place and, in spite of variations in I_c, to regard the value of ΔI_c resulting from a change of minority carrier concentration, as proportional to this change.

It should be noted that ΔI_c is the current increment for constant voltage across the barrier, and that gross changes in potential in the semiconductor on the application of emitter current must be compensated. Errors due to changes in floating potential, however, should now be negligible.

The choice of reverse voltage is governed by the requirements that it should be large enough to bias the contact into the saturation region, and small enough for drift velocities due to the field within the semiconductor to be small compared with diffusion velocities. The argument given here would not apply in the presence of current multiplication which is non-linear, nor, if it is desired to move the probe contact, in the presence of current multiplication which varies from contact to contact.

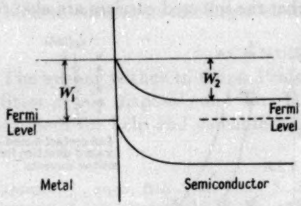


Fig. 1. Schematic diagram of electron energies at collector barrier, showing notation used in text.

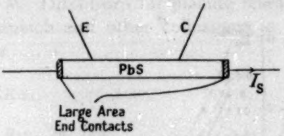


Fig. 2. Schematic diagram of arrangement for sweeping current experiments.

§ 2. EXPERIMENTAL METHOD

The above theory has been applied to the investigation of injection in p-type lead sulphide. In this material there is no evidence of variable current gain in a collector barrier. It was found possible to produce, by cleaving or by grinding, some single crystal 'filaments' of sizes of the order of 0·1 cm × 0·01 cm × 0·01 cm when starting from suitable crystals. These specimens were not of the highest purity (estimated carrier concentration at room temperature approaching $10^{17}/\text{cm}^3$), but injection phenomena could be obtained and studied in them. Attempts to obtain end electrodes of area greater than the cross section of the filament, as used by Shockley, Pearson and Haynes (1949), did not succeed.

The procedure adopted was as follows: (*a*) the point contact E (fig. 2) was applied to the filament, (*b*) the point C was applied at a known distance from E, and the floating potential v of C was measured for a fixed value of sweeping current I_s, but with no current through E, (*c*) a voltage $v+V$, where V was a small fixed positive voltage (e.g. +2 volts), was applied to C and I_c was measured, (*d*) a known forward current I_e was passed through contact E, (*e*) the new floating potential v' of the point C was measured, (*f*) a voltage $v'+V$ was applied to C, and the collector current $I_c' = I_c + \Delta I_c$ was measured.

These measurements were repeated for various (positive and negative) values of I_s. The point C was then moved to a new position, re-applied with care at a fixed pressure, and the operations were repeated. Measurements were also made in the absence of the emitter contact E, the point C being placed near one of the end contacts of the filament. In this case (*a*) with $I_s = 0$ the voltage V was applied to C and the current I_c was measured, (*b*) a known sweeping current I_s was passed and the floating potential v of the point C was measured, (*c*) a voltage $v+V$ was applied to C, and the current $I_c' = I_c + \Delta I_c$ was measured. These latter measurements were repeated for various values of I_s, and for various distances of C from the end contact.

§ 3. EXPERIMENTAL RESULTS

The results plotted in fig. 3 represent current increments ΔI_c which result from the passage of forward current through the neighbouring point contact E, plotted as a function of point separation. The curves show that the concentration of injected carriers decays rapidly with increasing distance from the emitter, and that the average drift distance is a function of the total sweeping current. Since I_e is comparable with I_s, the carriers are moving in a non-uniform field, and so calculations of lifetime from a mobility determined by Hall effect cannot be made. The polarity of the effect confirms that the injected carriers are electrons.

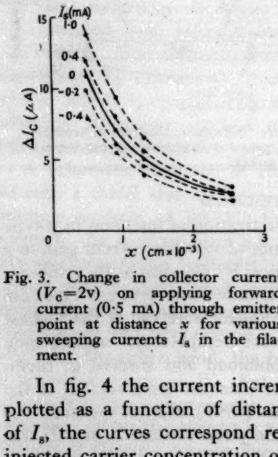


Fig. 3. Change in collector current (V_c=2v) on applying forward current (0·5 mA) through emitter point at distance x for various sweeping currents I_s in the filament.

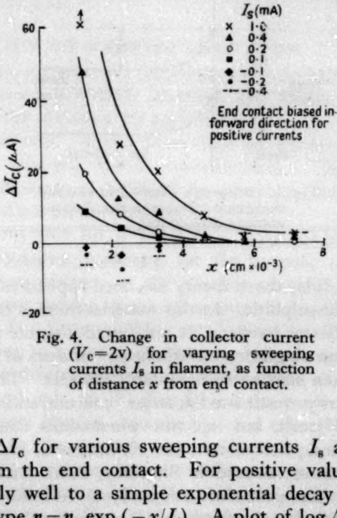


Fig. 4. Change in collector current (V_c=2v) for varying sweeping currents I_s in filament, as function of distance x from end contact.

In fig. 4 the current increments ΔI_c for various sweeping currents I_s are plotted as a function of distance from the end contact. For positive values of I_s, the curves correspond reasonably well to a simple exponential decay of injected carrier concentration of the type $n = n_0 \exp(-x/L)$. A plot of $\log \Delta I_c$ against x gives a set of parallel straight lines, the value of L being $1{\cdot}1 \times 10^{-3}$ cm. The fact that L is independent of I_s and so independent of the electric field in the crystal was found to be in agreement with predictions from potential measurements. The field in this region, for the largest current, was less than 10 v/cm, giving a potential change of less than 0·01 volt in a distance L of the order of 1×10^{-3} cm; this is appreciably less than $\boldsymbol{k}T/\boldsymbol{e} (=0{\cdot}025$ volts). Hence the diffusion process dominates the flow. Taking from Hall effect measurements a value of 100 cm^2 v^{-1} sec^{-1} for the mobility in this specimen, and calculating the diffusion constant D from the Einstein relationship, the expression $L = (D\tau)^{1/2}$ gives a value of order 5×10^{-7} sec for the lifetime τ of injected electrons.

It will also be seen that negative values of ΔI_c are obtained when the direction of the sweeping current I_s is reversed. It is the direction of this current which determines whether the end contact in question is biased in the forward or reverse direction. The positive increments of I_c show that electron injection takes place for forward bias voltages, whereas the negative changes of collector current must now be interpreted as a consequence of electron extraction, i.e. the removal by the end electrode from the semiconductor of free electrons which exist there by virtue of the thermal equilibrium between the full and the conduction bands. The apparently unsystematic variation of the negative values of ΔI_c may be due to the breakdown of the assumption that the local replacement rate for minority carriers is constant. A similar extraction effect has been observed in germanium.

The results of these experiments show that in these specimens of lead sulphide carrier injection occurs in the forward current from both 'point' and small area metal contacts, and there is also evidence of the opposite process, extraction, in the reverse current at the larger contacts.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. Ditchburn for placing research facilities at his disposal, and Dr. H. K. Henisch and other colleagues in the department for help and valuable suggestions.

REFERENCES

Bardeen, J., 1950, *Bell Syst. Tech. J.*, **29,** 469.
Bardeen, J., and Brattain, W. H., 1948, *Phys. Rev.*, **74,** 230.
Shockley, W., Pearson, G. L., and Haynes, J. R., 1949, *Bell Syst. Tech. J.*, **28,** 344.

A Precision Measurement of the ^{137}Cs γ-Line

By G. LINDSTRÖM, K. SIEGBAHN and A. H. WAPSTRA*

Nobel Institute for Physics, Stockholm, Sweden

Communicated by K. M. G. Siegbahn; *MS. received* 3*rd July* 1952

ABSTRACT. The recent establishment of a set of standard lines for precision work in β-ray spectroscopy is reviewed. In this work use was made of a semicircular spectrometer adjusted for high resolution, the proton resonance method for measuring the magnetic field and a large double focusing spectrometer. In this way absolute measurements of the lines were made, as well as a number of different relative measurements. The general consistency is very satisfactory, indicating a precision of the standard lines of 1 or 2 parts in 10^4. So far the following lines have been measured: the F, I, L and X lines of Th(B, C″), the 411 kev line of ^{198}Au, the annihilation radiation, the 2·75 Mev line of ^{24}Na, the 662 kev line of ^{137}Cs. In the present paper the measurements of the ^{137}Cs are described. The line has been studied as a conversion line and as a photo line emanating from a uranium converter. The combined results of the two spectrometers yield the following value of the energy of the ^{137}Cs line: $E_\gamma = 661{\cdot}65 \pm 0{\cdot}15$ kev. The $H\rho$ value of the conversion line is $3381{\cdot}28 \pm 0{\cdot}50$ gauss cm.

During the last two years, β-spectrometric methods of measuring the energies of γ-lines have been greatly improved, so that it is now possible to obtain energy values to within 1 or 2 parts in 10^4. So far the following set of standard lines has been measured: the F- and I-lines of ThB, the L- and X-lines of ThC″, the annihilation radiation, and the ^{198}Au γ-line. The absolute energy measurements of the Th lines were performed by Lindström (1951 a, b) using a semicircular spectrometer with a homogeneous field and the proton resonance absorption method for measuring the magnetic field. The resulting energy values were then based *either* on the very accurate value of the proton magnetic moment given by Thomas, Driscoll and Hipple (1950), *or* on known X-ray data. The mutual consistencies among the energy values thus obtained were very satisfactory. In particular cases in which there is no limit either to the total or to the specific activity of the radioactive source, this method seems to be the easiest to use, because of its simplicity and also because of the fact that the field can be measured in absolute units with great accuracy. Some difficulties are encountered when the magnetic resonance method is applied to measure the low fields connected with low energy lines.

When the total and specific activity is limited, or when a converter is being used for producing photo lines, extended sources have to be used which require large dispersion and thus, for convenience, double focusing. With a double focusing spectrometer having $\rho = 50$ cm, source areas of about 30 mm^2 can still be employed even when the resolution is set to be as good as about $1:10^3$. When this technique was applied to the measurement of the annihilation radiation (and other γ-lines) a mutual check was possible between the two methods, discussed above (Hedgran and Lind 1952). Thus it was found that complete agreement between the two different sets of measurements on the same line could be obtained only if the energy of the annihilation line was put exactly equal to m_0c^2, which can be calculated very accurately from known combinations of elementary constants.

---

* On leave from Instituut voor Kernphysisch Onderzoek, Amsterdam.

It seems very desirable to extend the above established set of precise standard lines to others, covering a larger energy region. In particular the ^{137}Cs γ-line at 0·66 Mev is extremely suitable. ^{137}Cs being a fission product, it can be prepared with very high total and specific activity. It has a long half life (37 years) and the 0·66 transition (being of the M4 type) is highly converted ($\alpha_K = 0{\cdot}095$). Since the main continuous β-spectrum (95%) has an upper limit of only 0·51 Mev the strong conversion line falls outside this spectrum.

A precise measurement of the energy of the ^{137}Cs γ-line has been made before by Langer and Moffat (1950). They compared the ^{137}Cs conversion line with that of ^{198}Au. The relative half-widths of the lines were about 0·5–0·7%. Their final value, $E_\gamma = 0{\cdot}6614$ Mev, was based on the crystal value of the ^{198}Au radiation as determined by DuMond *et al.* (1948, 1949). In establishing the new set of standard lines described above it was found, however, that the crystal value was $1:10^3$ too low, making a correction of about this magnitude necessary in Langer and Moffat's value for the ^{137}Cs line.

In the present investigation we have used the double focusing spectrometer as well as the homogeneous field spectrometer with the proton resonance equipment for measuring the magnetic field. In order to eliminate a number of possible sources of systematic error, several more or less independent ways of comparing the energy of the ^{137}Cs γ-line with reference energies have been used. It will be shown below that these measurements are all consistent to a very high degree with each other, so that the final energy value obtained as a result of the combined measurements can be given with some confidence.

The measurements with the double focusing spectrometer will be discussed first. Using a thin evaporated uranium converter (size 4 mm × 15 mm, thickness 0·7 mg/cm^2) in a standard position, the ^{137}Cs K photo line was compared to the K photo line of the annihilation radiation expelled from the same converter. With the resolution used in these experiments the Doppler broadening of the annihilation line is quite pronounced. This effect has been investigated in the spectrometer before and is described in a paper by Lind and Hedgran (1952). In order to get a precise energy comparison between the photo lines of different widths the folding procedure described in their paper was used. In calculating the energy of the Cs line the energy of the annihilation radiation was assumed to be exactly equal to m_0c^2. The theoretical reason for this is quite strong and has furthermore been substantiated by the measurements mentioned earlier in this paper. According to the recent least squares adjustment of the atomic constants by DuMond and Cohen (1951) $m_0c^2 = 510{\cdot}969 \pm 0{\cdot}010$ kev.

Using the same uranium converter the K photo line of ^{137}Cs was also compared to the K photo line of the 510·85 kev γ-radiation of the ThC″ (the 'L' line in Ellis' notation). The K conversion line of this transition has been measured absolutely by Lindström, and the energy ratio between the 510·85 kev line and the annihilation radiation (as K photo lines from uranium) has been measured by Hedgran and Lind. Lindström's absolute measurements gave as a result $E_L = 510{\cdot}85 \pm 0{\cdot}08$ kev. Taking the energy of the annihilation radiation as $m_0c^2 = 510{\cdot}969$, the relative measurements by Hedgran and Lind result in $E_L = 510{\cdot}83 \pm 0{\cdot}14$ kev. The relative measurement in the present investigation between the ^{137}Cs photo line and the ThC″ 510 kev photo line will then give us an energy value of the ^{137}Cs line which should agree with the value obtained from the first independent relative measurements described above of the same line with respect to the annihilation radiation, i.e. the energies of the three lines form a closed cycle, which is very valuable as a check of the internal consistencies.

The energy comparisons between the uranium photo lines have been performed by fitting the high energy slopes of the lines to each other in order to eliminate possible straggling effects in the converter. At the electron energies concerned here, however, and for the uranium converter used, it can be calculated that the probability even for single electron scattering inside the converter is fairly small. Experimentally it was found that the line-shapes of, for example, the 510 kev ThC″ photo line agreed quite well with that of the ^{137}Cs photo line.

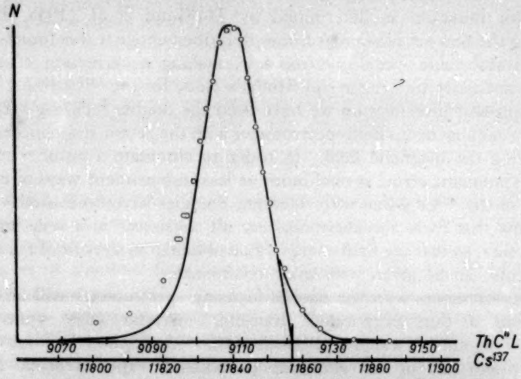


Fig. 1. An energy comparison between the ^{137}Cs conversion line (circles) and the ThC″ 510 kev line (full curve).

The last series of energy measurements were performed by comparing the internal conversion line of the ^{137}Cs line with the 'L' internal conversion line in the ThC″ spectrum. The $H\rho$ value of the latter line is $2607{\cdot}17 \pm 0{\cdot}30$, according to Lindström's absolute $H\rho$ measurements.

The converter effects are eliminated in this case. One other uncertainty is simultaneously introduced however: the non-uniformity of the electron-emitting ^{137}Cs surface may limit the exact definition of the source position. Because of the field form and the large dispersion of the instrument a shift of 0·2 mm means an error of only $1:10^4$ in the $H\rho$ value.

In order to get a uniform distribution of the Cs activity the insulin method of Langer (1949) was used. The activity was deposited on an aluminium foil with thickness 0·2 mg/cm^2, on an area of 4×18 mm^2. One may suspect a certain concentration of the activity at the edges of the rectangular area. A slit 2 mm × 13 mm placed on the aluminium foil masked off the dangerous regions. The same slit was used for the active Th source, which is very uniform, being prepared in an activation vessel using a RdTh sample.

Figure 1 shows one energy comparison between the ^{137}Cs conversion line (circles) and the ThC″ 510 kev line (full curve). The abscissae are fitted so that the high energy slopes coincide. In this run the relative line-width was set to be $1{\cdot}6 \times 10^{-3}$. At this resolution the finite thickness of the Cs source shows up as a small broadening at the low energy side of the line. A very close fit can, however, be made on the high energy sides. Four independent runs were made on the conversion line comparison.

Before presenting the results of the above energy measurements on the ^{137}Cs γ-radiation by means of the double focusing spectrometer, a possible source of systematic error should be discussed. The magnetic field is always measured at one fixed point and it is assumed that the average field along the electron paths between different field settings varies in exactly the same way as does the field at the fixed point. Since the field is shaped by means of iron one cannot expect that this assumption is justified to an arbitrarily high degree of accuracy. There are in fact strong indications that this effect sets an ultimate limit to the highest precision which can be attained with this type of instrument.*

If a close doublet is examined, for example the U_K photo line of the annihilation radiation and the $U_{L_{III}}$ photo line of ^{198}Au (411 kev radiation), the above effect needs no consideration. In the present investigation, however, the ^{137}Cs line has been compared with other lines, differing by approximately 30% in $H\rho$. In order to check the linearity of the instrument over large $H\rho$ intervals an extensive series of relative measurements was made on the F, I and L conversion lines in the spectrum ThB + C″. The results were compared with those obtained by Lindström when he studied the same lines using the proton resonance method and the semicircular spectrometer. It was found that individual relative measurements between two lines showed small jumps which were outside the limits which could be accounted for by statistics and the measuring procedure alone. The *mean* ratios of I/F and L/I agreed, however, within about ±1 in 10^4 with those found by Lindström (1951). Later measurements in the region of very high energies (~2·5 MeV) seem to indicate a definite non-linearity of the instrument (Hedgran and Lind 1952). The effect is, however, just too small to be considered in the present investigation, which claims an accuracy within 2 in 10^4.

In table 1 we give the final results of the measurements made with the double focusing spectrometer.

Table 1

Reference line	Cs-line	$H\rho$ of Cs-line (gauss (cm)	Kinetic energy (kev)	Binding energy (kev)	E_γ
(Th–L)	(^{137}Cs K-conv.)	3381·75 ± 0·70	624·33 ± 0·20	37·44	661·77 ± 0·20
(ThC″U_K)	(^{137}Cs U_K)	3083·86 ± 1·0	545·35 ± 0·30	115·59	661·94 ± 0·30
($e^+e^-U_K$)	(^{137}Cs U_K)	3083·22 ± 1·0	545·18 ± 0·30	115·59	661·77 ± 0·30

The binding energies are calculated from the corresponding L_{III} edges by using the constants given by DuMond and Cohen (1951), and the X-ray data given in Cauchois and Hulubei's tables (1947).†

---

* An *iron-free* double focusing spectrometer is not limited in precision by this effect. Such an instrument is now under construction.

† The fact that the binding energies obtained from X-ray absorption data do not correspond to the actual work of liberating an electron from the source does not cause any further uncertainty in the energy value at the γ energy concerned here, and with the limits of error given here.

As mentioned earlier in this paper we have also determined the *Hρ*-value of the ^{137}Cs K-conversion line in absolute units by using the small semicircular β-spectrograph and the proton resonance probe.

The difficulty in this case is to prepare a thin source with high activity and small dimensions. The source used consists of activated wires with a diameter ranging from 25 μ to 35 μ. Two Geiger–Müller tubes in coincidence were used as detectors in this case. The distance between the Geiger–Müller slit and the source (2ρ) was of the order of 95 mm.

The sources were prepared by alternately placing the wires in a $PtCl_4$ solution and in a concentrated KCl solution. The purpose of this is to deposit non-soluble K_2PtCl_6 on the wire. This procedure was repeated four or five times. The wire was then activated by placing a small part (about 3 mm) of it in the active Cs solution for about 24 hours. There will then be an exchange between K^+ and Cs^+ ions.

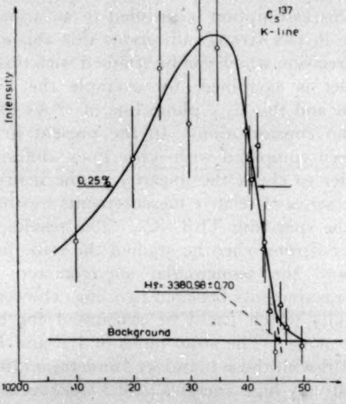


Fig. 2. An experimental curve taken with the semicircular β-spectrograph, indicating how the *Hρ*-value was measured.

The individual results are tabulated in table 2.

Table 2

Run	1	2	3	4
Hρ (gauss cm*)	3381·17	3380·34	3380·98	3380·76
Source	A	B	C	C

Mean value 3380·81 ± 0·70. $E = 661{\cdot}54 \pm 0{\cdot}20$ kev.

A Pt wire, with diameter 25μ, B tungsten wire, diameter 36μ, C tungsten wire, diameter 26μ.

* The *Hρ* values are calculated by using $\gamma_p = (2{\cdot}67523 \pm 0{\cdot}00006) \times 10^4$ sec^{-1} gauss^{-1}.

Three different wires were prepared to check whether the activity was evenly distributed over the surface of the wire. The agreement between the values of table 2 indicates that this is the case.

The field was measured along the path with a small proton sample, at the field corresponding to the high energy side of the line. This part of the line does not seem to be disturbed by the source thickness.

The ρ-values have been measured by a comparator. 2ρ then corresponds to the distance from the nearest slit edge to the nearest edge of the wire.

The sources were rather weak, which means that the statistical deviation in the individual points were rather large, as may be seen in fig. 2. This is the reason why the estimated error was taken to be about ± 2 in 10^4 in spite of the fact that better agreement between the individual values was actually obtained.

RESULTS

The uncertainties in the different measurements described in this paper are of the same order of magnitude, and the consistency indicates that the following mean values will be correct within the given limits of error: $H\rho$ (^{137}Cs K-conv. line) $3381{\cdot}28 \pm 0{\cdot}50$ gauss cm, E_γ (^{137}Cs) $661{\cdot}65 \pm 0{\cdot}15$ kev. These values are compared with other reference lines measured at this laboratory in table 3.

Table 3

Line	$H\rho$ (gauss cm)	E_γ (kev)	Method	Reference
Th-F	$1388{\cdot}56 \pm 0{\cdot}15$*	$238{\cdot}63 \pm 0{\cdot}04$	absolute	Lindström (1951 a, b)
Th-I	$1754{\cdot}01 \pm 0{\cdot}20$	$238{\cdot}62 \pm 0{\cdot}04$	absolute	
Th-L	$2607{\cdot}17 \pm 0{\cdot}30$	$510{\cdot}85 \pm 0{\cdot}08$	absolute	
^{198}Au K-conv.	$2222{\cdot}4 \pm 0{\cdot}4$	$411{\cdot}75 \pm 0{\cdot}10$	against Th-L	Lind and Hedgran (1952)
^{137}Cs K-conv.	$3381{\cdot}28 \pm 0{\cdot}5$	$661{\cdot}65 \pm 0{\cdot}15$	absolute, against Th-L, against annihilation	Present investigation
Th-X	$9986{\cdot}7 \pm 1{\cdot}5$†	$2614{\cdot}25 \pm 0{\cdot}50$	absolute	Lindström (1952)
^{24}Na U_K	$10363{\cdot}5 \pm 4$	$2753{\cdot}5 \pm 1{\cdot}0$	against Th-X	Hedgran and Lind (1952)

* These values have later been confirmed by Craig (1952). He gives $1388{\cdot}5 \pm 0{\cdot}3$ and $1753{\cdot}9 \pm 0{\cdot}4$ respectively.

† This line has also been measured by Brown (1951). His value is 9988 ± 2.

The ratios between the different lines given in the table have also been measured. The measurements prove that the consistency between the different values is well inside the given limits of error.

REFERENCES

Brown, W. L., 1951, *Phys. Rev.*, **83,** 271.

Cauchois, Y., and Hulubei, H., 1947, *Longueurs d'Onde des Emissions X et des discontinuités d'absorption X* (Paris: Hermann).

Craig, H., 1952, *Phys. Rev.*, **85,** 688.

DuMond, J. W. M., and Cohen, E. R., 1951, *Phys. Rev.*, **82,** 555.

DuMond, J. W. M., Lind, D. A., and Watson, B. B., 1948, *Phys. Rev.*, **73,** 1392; 1949, *Ibid.*, **75,** 1226.

Hedgran, A., and Lind, D. A., 1952, *Ark. Fys.*, **5,** 177.

Langer, L., 1949, *Rev. Sci. Instrum.*, **20,** 216.

Langer, L., and Moffat, R., 1950, *Phys. Rev.*, **78,** 74.

Lind, D. A., Brown, J. R., and DuMond, J. W. M., 1949, *Phys. Rev.*, **76,** 1838.

Lind, D. A., and Hedgran, A., 1952, *Ark. Fys.*, **5,** 29.

Lindström, G., 1951 a, *Phys. Rev.*, **83,** 465; 1951 b, *Ark. Fys.*, **4,** 1; 1952, *Phys. Rev.*, **87,** 678.

Thomas, H. A., Driscoll, D. L., and Hipple, J. A., 1950, *Phys. Rev.*, **78,** 787.

LETTER TO THE EDITOR

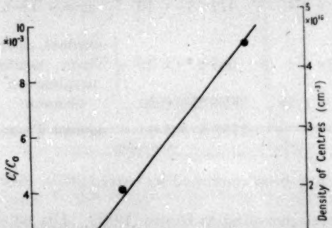
Distribution Coefficient of Indium in Germanium on Crystallization

The distribution of indium impurity in a germanium ingot has been measured using the radioactive tracer technique. The germanium used was electrically n-type with a peak reverse voltage of 50–70 volts. The radioactive indium was prepared from metal of 99·9% purity, activated by neutron bombardment at the Atomic Energy Research Establishment, Harwell.

The ingot was grown by the method of Teal and Little (1950) in an atmosphere of argon from a 14 g germanium melt containing $2{\cdot}3\times10^{-3}$ g radioactive indium. During growth the melt was stirred fairly vigorously by jets of argon directed at the surface, the direction of rotation being periodically reversed. The rate of growth was about 2·5 cm per hour.

The resulting ingot was a single crystal approximately 4 cm long and 0·8 cm average diameter. It was electrically p-type throughout.

The ingot was sectioned at right angles to its length and the sections weighed and etched. The activity of each section was then measured with an end-window Geiger–Müller counter, using a standard lead castle. The counting rates (corrected for the area and thickness of each specimen) determine the relative values of indium concentration in the various sections; to determine their absolute magnitude one section was dissolved in *aqua regia*, made up to a standard volume, and the activity of the solution compared, using a liquid counter, with that of a standard solution containing a known concentration of the radioactive indium.



The distribution of the concentration of impurity in the ingot should obey the relation (Pfann 1952 a)

$$C/C_0 = k(1-g)^{k-1} \qquad \ldots\ldots (1\,a)$$

$$\simeq k/(1-g) \qquad (k \ll 1), \qquad \ldots\ldots (1\,b)$$

provided that mixing in the liquid phase is complete, and that diffusion in the solid phase is negligible; in this equation C and C_0 are respectively the impurity concentrations (in mass units) in the ingot and in the initial melt, g is the fraction of the melt which has solidified, and k is the distribution coefficient, i.e. the ratio of impurity concentration in solid and liquid on either side of the growing interface under equilibrium conditions. The measured concentrations are shown plotted in the figure as a function of $(1-g)^{-1}$; relation

(1 *b*) is seen to be obeyed fairly well. The distribution coefficient *k* calculated from the slope of the graph, is $(1{\cdot}25 \pm 0{\cdot}15) \times 10^{-3}$. This is in satisfactory agreement with an approximate value of 1×10^{-3} quoted by Pfann (1952 b) from measurements made at the Bell Telephone Laboratories.

Research Laboratory,
Associated Electrical Industries Ltd.,
Aldermaston, Berkshire.
15th October 1952.

J. J. DOWD.
R. L. ROUSE.

PFANN, W. G., 1952 a, *J. Metals*, **4**, 747; 1952 b, *Ibid.*, **4**, 861.
TEAL, G. K., and LITTLE, J. B., 1950, *Phys. Rev.*, **78**, 647.

REVIEWS OF BOOKS

Empfangsprobleme im Ultrahochfrequenzgebiet, by H. F. MATARÉ. Pp. 264. (Munich: R. Oldenbourg, 1951.) DM 33.

In this book the author assumes that the reader is familiar with classical receiver theory, as applied to wavelengths greater than some 50 cm, and gives his attention to the special problems which arise in the decimetre wave and microwave regions. From the author's point of view a main characteristic of the long wave region ($\lambda > 50$ cm) is that the noise arising from external sources normally exceeds the noise generated inside the receiver itself. In the decimetre wave and centimetre wave ranges, however, noise generated internally in the stages preceding the intermediate-frequency amplifier is the decisive factor limiting receiver performance. Since the faintest usable signal is that which just exceeds the noise level, the book is largely concerned with the study of noise in the first amplifier and mixer or rectifier stages.

In introductory chapters an outline is given of the nature and scope of the problems to be discussed, including a statement which defines the purpose of the work " es ist stets das Ziel, vorauszuberechnen und dann zu bauen ". But, although the appeal is perhaps chiefly to the high-frequency engineer, the physical problems are of great interest and the author's coherent treatment of material at present scattered through an extensive literature is very attractive. Some hundred references are given to the chief sources, but the book is not a compendium built up from the literature. The author has worked for ten years in this field and the material presented is so integrated that the reader is conscious of a developing theme as the book proceeds.

A general discussion on reception at ultra-high frequencies is followed by a statement of the methods adopted in sensitivity calculations which provides a key to the subsequent chapters on diode mixing, triode mixing and high-frequency amplification, all applied to decimetre waves. This is followed by a study of centimetre wave receiver problems including interesting chapters on crystal rectifiers, semiconductors and waveguides. A section on measurement problems concludes the work.

The emphasis of the book as a whole is utilitarian. In each section the author makes his intention clear, giving an outline of the special conditions to be considered and of the method to be adopted in handling the problem before developing the appropriate equations. Possible simplifications are indicated in the text and the meaning of the various expressions is brought out by means of excellent graphs. These are invariably provided with numerical coordinate scales, so that the designer can answer his own questions from the graphs without the necessity for elaborate calculation. Worked examples are given of the use of these graphs, as well as of the equations, in elucidating the performance of various proposed circuits; in many cases these are followed by accounts of the author's own measurements on the types of circuit considered. Except for the worked examples and descriptions of measurement, not much practical detail is presented.

The book is not easy to read; the material presented is of a high standard and much of the argument is essentially mathematical. It is a valuable reference work for anyone interested in ultra-high-frequency measurements and communications. W. A. PROWSE.

Acoustics in Modern Building Practice, by F. INGERSLEV. Pp. x + 290. (London: The Architectural Press, 1952.) 35*s*.

Among the well-known laboratories for building acoustics, that at the Danish Academy of Technical Sciences, Copenhagen, though among the youngest, is not the least famous. The author of this book is its Director; he here expounds the science as practised in his Institute and, though much of the technique is standard in acoustic laboratories, is able to show new applications to building science, a science in which the Scandinavian countries show great originality and, indeed, seem to be ahead of most countries (as Mr. W. A. Allen points out in the foreword).

In other respects the book is rather reminiscent of similar textbooks (some familiar Sabine diagrams and N.P.L. photographs will be spotted by the reader), and there is a tendency, which some physicists abhor, to group a few pieces of equipment in boxes to form a photograph without explaining what is inside ; but the photographs of special acoustical treatment of Danish interiors are well worth showing and not readily available elsewhere.

The book is thoroughly up-to-date in its treatment of sound absorbent theory without being too mathematical, and always with consideration to practical application. Chapters on noise suppression are included.

In short, the book is to be recommended to the practitioner in technical acoustics who wants an up-to-date survey of architectural acoustics. E. G. R.

Text Book on Sound, by J. W. WINSTANLEY. Pp. xi + 239. (London: Longmans Green, 1952.) 12*s*. 6*d*.

The author is Physics Master at Manchester Grammar School and his book was written for the sixth form of schools and the first year at the University with the hope of making the senior pupil acquainted with this ' Cinderella of the Sciences ', as he calls acoustics.

The book follows the usual course, beginning with something about sound propagation and the simple mathematics of vibrating systems and leading up to the various types of the latter, except that, rather curiously, the compounding of S.H.M.'s is left to the end. Applications are dealt with in two chapters under ' Acoustics of buildings ' and ' Analysis, recording and reproduction of sound '. The book is quite up to date and the student's interest sustained, though one wonders if he is not expected to be able to read fairly involved valve circuits at an early stage. But probably the modern schoolboy is more adept at this than in appreciating the fundamentals of physics.

Very useful to both student and teacher are the questions culled from many examinations and preceded by worked examples. Some of these follow each chapter. E. G. R.

Le Contraste de phase en optique et en microscopie, by MAURICE FRANÇON. Pp. 108. (Paris: Editions de la Revue d'Optique Théorique et Instrumentale, 1950.) 480 fr.

This work is primarily an essay on the general mathematical background underlying procedures analogous to (and including) the phase contrast method of Zernike, supplemented by a description of the author's own phase contrast ' converters ', which enable the method to be used with an ordinary micro-objective. It is not a textbook, nor a balanced survey of the subject, nor does it contain very much which is likely to be of service to the practical user of phase contrast microscopes, but it has an interest of its own which entitles it to a place in the optical literature. E. H. LINFOOT.

Technical Optics, Vol. II, by L. C. MARTIN. Pp. 344. (London: Pitman, 1950.) 40*s*.

For many years Martin's *Applied Optics* Vols. I and II have been standard text books of general optics up to degree standard, and have been the starting point of many who have studied the subject further. Most optical research workers keep copies on the shelf for occasional reference. A revised and enlarged edition is now published under this modified title; as previously, Volume II is concerned with optical instruments, their underlying principles and methods of testing, whilst Volume I deals mainly with fundamental theory.

The first chapter is unchanged, whilst to the second has been added an account of modern eyepieces, periscopic and flexible optical systems and sighting telescopes. In Chapter III instructions for setting-up the microscope have been omitted, and a useful treatment of phase contrast included. The section on binocular instruments has been enlarged by an account of grid, lenticular grid and vectograph methods of stereoscopic projection and of modern theory of stereoscopic vision. The arrangement of the material of Chapter V (Photographic Lenses) has been improved and the section dealing with the correction of the off-axis image field aberrations enlarged. The treatment of the photometry of optical systems has been brought more up-to-date, and now includes a description of the Stiles-Crawford Effect and of the measurement of the light transmission of optical instruments using photoelectric devices. Chapter VII (The Testing of Optical Instruments) has been enlarged considerably; additions include accounts of the measurement of resolution and contrast reduction and a description of the Platzeck and Gaviola Test.

A very interesting chapter has been added, headed Aspheric Surfaces of Revolution. This deals with the application of such surfaces (for example, to reflecting microscope objectives) and gives a condensed account of various methods of producing them.

Both type and diagrams are clear, and references to other publications are plentiful. Those who are familiar with its predecessor will need little persuasion to obtain this volume for the additions it contains, and others are fortunate that the modernized edition is available. K. J. HABELL.

Music and Sound, by LL. S. LLOYD. Pp. xiv+181. 2nd edn. (Oxford: University Press, 1951.) 16*s*.

This is the second edition of a book which is addressed to students of music. There is, however, a sufficient body of physicists with musical leanings to justify a review of the book in a physical journal. A certain amount of musical knowledge is assumed, but the book could be read intelligently by a physicist who knows only those parts of the theory of music in the syllabuses of the Associated Board examinations.

The pure scale is first discussed and the use of logarithms painlessly introduced. The troublesome ‘ comma ’ is unearthed, and then the use of the mean-tone and equal temperaments is introduced, and the possibilities of modulation explained.

Chapters IV and V introduce the reader to the name of Helmholtz, combination tones and the physics of harmony. The indiscretions of physicists in providing ‘ theories ’ of harmony are treated, when at all, very clemently.

The remaining chapters deal with the quality of musical sounds, properties of the ear and of wind and stringed instruments.

For students of music the book represents a sound account of the physics they need. Many physicists will find it both readable and valuable. H. H. HOPKINS.

CORRIGENDA

Photoelectric Analysis of Elliptically Polarized Light, by J. F. ARCHARD, P. L. CLEGG and A. M. TAYLOR (*Proc. Phys. Soc.* B, 1952, **65**, 758).

P. 762. In the table showing effects of misadjustments the stated orientations of the ellipse for misadjustment of the polarizer and angle of incidence should be interchanged. The subsequent discussion is not affected.

The Optical Constants of Thin Metallic Films Deposited by Evaporation, by P. L. CLEGG (*Proc. Phys. Soc.* B, 1952, **65**, 774).

P. 781. The reference to Avery, D. G., *should read* 1950, *Phil. Mag.*, **41,** 1018.

CONTENTS OF SECTION A

PAGE

Editorial 1

Dr. G. R. Baldock. Determination of the Surface Energy of a Metal by Molecular Orbitals 2

Miss C. F. Lees, Mr. G. C. Morrison and Dr. W. G. V. Rosser. The Range–Energy Relation for Protons and Alpha-Particles in Diluted Ilford G5 Emulsions 13

Dr. J. Irving. The Effect of the Tensor Force on the Binding Energy of the Alpha-Particle 17

Dr. R. H. Dalitz. Some Features of the Deuteron Stripping Process . . . 28

Dr. G. R. Heyland and Dr. W. E. Duncanson. A Search for Irregularities in the Absorption of Cosmic Rays in Lead 33

Dr. G. R. Heyland and Dr. W. E. Duncanson. Momentum Distribution for Cosmic Ray Mesons up to 6 kmev/c 40

Dr. A. L. Hodson. Some Aspects of the Altitude Variation of Extensive Air Showers 49

Dr. A. L. Hodson. Penetrating Particles in Extensive Air Showers . . . 65

Dr. E. Robinson. Spark Counters for Short Time Interval Cosmic Ray Measurements 73

Dr. E. Robinson. Short Time Interval Measurements on Pairs of Associated Cosmic Ray Particles 79

Mr. E. F. W. Seymour. Nuclear Magnetic Resonance Line Width Transition in Aluminium 85

Dr. J. M. Ziman. Antiferromagnetism by the Spin Wave Method—III: Application to more Complex Systems 89

Dr. R. Middleton, Dr. F. A. El-Bedewi and Dr. C. T. Tai. An Investigation of the Neutron Groups from the Reactions ^{12}C(d, n)^{13}N, ^{16}O (d, n) ^{17}F and ^{32}S (d, n) ^{33}Cl 95

Dr. W. H. Evans, Mr. T. S. Green and Dr. R. Middleton. An Investigation of the Reaction ^{14}N(d, n)^{15}O at 8 mev Deuteron Energy 108

Research Note :

Dr. A. J. Herz and Miss M. Edgar. The Reduction of Distortion in Nuclear-Research Emulsion 115

Letters to the Editor :

Prof. L. Jánossy. Note on the Fluctuation Problem of Cascades . . . 117

Mr. R. S. Trenam. Paramagnetic Resonance in Four Double Nitrate Salts 118

Mr. B. Rose and Dr. J. M. Freeman. Inelastic Neutron Scattering in Iron 120

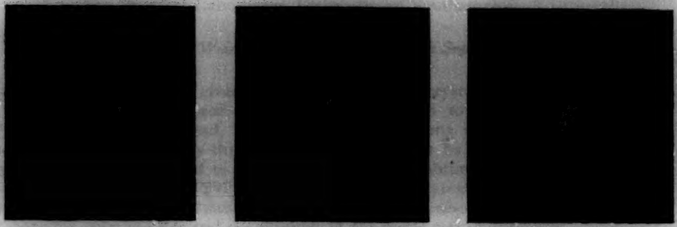
Prof. E. A. Guggenheim. Electric Spin in Semiconductors . . . 121

Mr. G. A. Erskine and Dr. M. J. Seaton. The Asymptotic Solution of an Equation Occurring in Scattering Theory 123

Reviews of Books 124

Contents of Section B 128

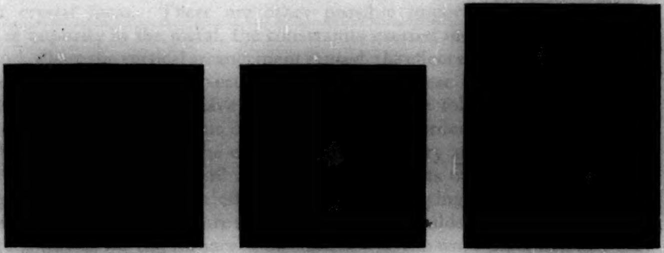
PHYS. PROC. SOC. VOL. 66, PT. I—B (D. FISHER)



(*a*) (*b*) (*c*)

Fig. 1. Diffraction patterns from β gutta percha.

(*a*) Unstretched specimen viewed perpendicular to electron beam.
(*b*) Unstretched specimen inclined at 45° to beam.
(*c*) Stretched specimen (300% elongation) viewed perpendicular to beam, showing traces of α structure.



(*a*) (*b*) (*c*)

Fig. 2. Diffraction patterns showing change in structure on stretching γ gutta percha.

(*a*) Unstretched specimen viewed perpendicular to beam—γ structure.
(*b*) Unstretched specimen inclined at 45° to beam—γ structure.
(*c*) Stretched specimen (300% elongation) viewed perpendicular to beam—mixed α and β structures.

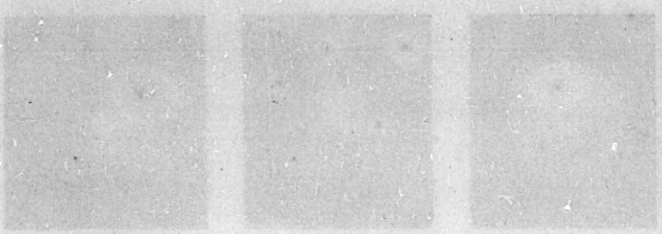


Fig. 1. Diffraction patterns from [illegible]

(a) Unstretched specimen [illegible]

(b) Unstretched specimen [illegible]

(c) Stretched specimen [illegible] showing traces of [illegible]

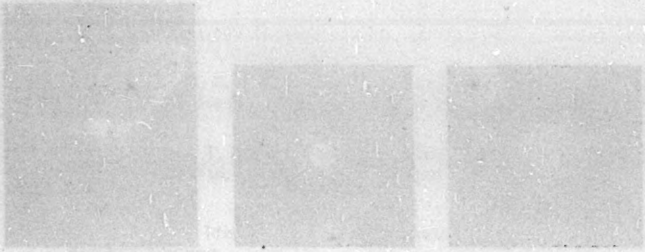


Fig. 2. Diffraction patterns showing change in structure on [illegible]

(a) Unstretched specimen [illegible]

(b) Unstretched specimen [illegible]

(c) Stretched specimen [illegible] perpendicular to [illegible] structure.

The Effect of Added Metallic Impurity and of Closed Moulds on the Growth from the Melt of Single Crystals of Tin*

By A. J. GOSS
The Physical Laboratory, University of Southampton

Communicated by S. Weintroub; *MS. received* 29*th September* 1952

ABSTRACT. A description is given of experiments on the crystallization of binary alloys of tin. The alloys were prepared from 99·997% pure tin to which known amounts (0·01–1%) of cadmium, zinc, lead, silver, indium or antimony were added. The effect of the added metal, as judged by the prevention of growth or by lack of perfection of the monocrystalline form, was found to depend on its solid solubility. Further experiments on the crystallization of the 99·997% pure tin in closed Pyrex moulds, in which both horizontal and vertical travelling furnaces were used, are described. The results were the same as those for open moulds and the position of the metal was unimportant.

§ 1. INTRODUCTION

In a previous communication (Goss and Weintroub 1952, referred to as I) experiments were described in which the growth of single crystals of tin, lead, zinc and bismuth, by the method of slow solidification from the melt was studied. It was shown that the rate of growth and thus the form of the temperature gradient in the metal was an extremely important factor governing the resultant crystal form. There are other possible important factors, e.g. the amount of impurity in the metal, the constraints exerted on the metal due to the mould and, where the vertical arrangement is used, the effect of the superincumbent liquid metal. No previous investigations in which these conditions have been examined in detail appear to have been made. In the following a description is given of a series of experiments on tin carried out in order to test the influence of the above factors. In §2 the crystallization of binary alloys of tin consisting of 99·997% pure tin† with small added known amounts of cadmium, zinc, lead silver, indium or antimony, briefly referred to as alloying metals, is dealt with and in §4 the crystallization of pure tin in closed Pyrex moulds using both horizontal and vertical travels for the furnace is discussed. In the work described in I only open end Pyrex moulds and horizontal travel were used. The results obtained in the present investigation indicate that the crystallization depends in all cases on the solid solubility of the impurity and that the effect is more marked at fast than at slow rates of growth. Neither the mould nor the weight of superincumbent liquid metal had any marked effect on the crystallization of tin.

§ 2. THE CRYSTALLIZATION OF BINARY ALLOYS OF TIN

Each alloy consisted of pure (99·997%) tin to which a known amount (0·01–1% by weight) of cadmium, zinc, lead, silver, indium or antimony had been added. A 500 g block of tin was weighed to the nearest 0·2 g and the appropriate quantity of clean pure alloying metal to the nearest 0·001 g. These were melted together *in vacuo* in a Pyrex tube the mixture being shaken continuously for about 10 minutes. When cool the alloy block, 25 mm in diameter, was removed from the tube and re-cast *in vacuo* into a long rod of 12 mm diameter. The ends of this rod were examined spectrographically to check the distribution of the alloying metal in the tin. Finally 12 specimen rods each 5 mm in diameter and 30 cm in length were cast. Rods of some of the mixtures had surfaces on which imperfections of the type shown in fig. 1 (Plate) could be seen. This kind of surface is briefly referred to as a non-specular reflecting surface. It is interesting to note that an empirical method of judging the purity of cast tin by visual examination of the surface has long been used by foundrymen.

---

* This work forms part of a thesis accepted by the University of London for the degree of Ph.D.
† The author is indebted to Messrs. Capper Pass & Co. for presenting the metal.

The specimens were crystallized in groups of four, using a horizontal travelling furnace and 7 mm bore open-end cylindrical Pyrex moulds. Three rates of growth, 21, 5 and 0·5 mm/min were used with a temperature gradient of the order of 20° c/cm. The temperature gradient was obtained from readings of a chromel–alumel thermocouple embedded in one of the specimens.

The form of the crystals was examined and the crystal orientations measured using the special form of optical goniometer (Goss 1952). The impurity metals were chosen to provide examples of various crystal structures and a range of solid solubilities in tin. Each batch of 12 specimens of a particular alloy was crystallized twice.

(i) *The Spectrographic Tests*

The specimens were examined spectrographically in the way advocated by Smith (1948). In general, close agreement with Smith's observations was obtained. Using a Hilger E2 medium quartz spectrograph the spectra covering the range 2100–6000 Å were recorded photographically. The source was a condensed spark and the addition of an inductance to the circuit, as recommended by Smith, reduced the intensity of the background and also of many of the longer wavelength lines with consequent improvement in the photographic record.

The electrodes were tin rods, the upper of pure tin and the lower of the alloy under test. The vertical position of the source relative to the slit was found to be critical.

The internal standard method of spectrographic analysis, in which the intensities of selected lines of the alloying metals are compared visually with the intensities of certain adjacent tin lines, was chosen. The accuracy varied with the various alloying metals; in general, however, a difference of 20% of alloying metal could be resolved. In almost every instance agreement was obtained with Smith (1948) which is considered as satisfactory.

The etched crystals were tested spectrographically at both ends. Any tendency for the alloying metal to accumulate towards the end of a crystal was noticed and, at the same time, the spectra of the twelve crystals containing various concentrations of a particular alloying metal were compared. The spectrographic examination was limited to an approximately circular area of 4 mm diameter. Some comparison was made between the spectra of crystals with surfaces in the three conditions: (*a*) unetched, (*b*) etched, and (*c*) clean, obtained by cutting approximately 1 mm from their ends. The records for (*b*) and (*c*) agreed and were taken as representative of the distributed metal in the tin whereas records for (*a*) showed, for some metals, a concentration of alloying metal on the surface.

§3. EXPERIMENTAL RESULTS

In the following sections observations of interest made during the preparation and crystallization of the binary alloys are listed. The orientations of the crystals and the temperatures in the crystallizing metals did not differ from those obtained previously for pure tin (I, pp. 566, 568) and are not therefore discussed further. Reference was made to a publication of the Tin Research Institute (1949) for information on the binary diagrams.

(i) *Tin–Cadmium*

Cadmium (hexagonal structure, $c/a = 1{\cdot}886$) has a solid-solubility limit in tin of 1% and because of the considerable accuracy with which this limit has been fixed cadmium was the first alloying metal investigated. Two dilute binary alloys containing 0·1 and 0·01% of cadmium were prepared from 99·999% pure cadmium, kindly given by the Imperial Smelting Corporation, and 99·997% pure tin. During the casting of the alloy rods, thin films of cadmium were deposited on the inside of the Pyrex envelopes. For the 0·1% alloy the weight of cadmium lost in this way was negligible; for the 0·01% alloy the slight loss was detectable. The cadmium dissolved easily in the tin and was found spectrographically to be evenly distributed. The alloy rods did not adhere to the Pyrex but the rod surfaces, particularly for the 0·1% alloy, showed the non-specular type reflection (fig. 1).

For pure tin the crystalline forms were described in I (p. 562). For the tin–cadmium 0·1% alloy the essential differences were that the first one or two centimetres of the rod consisted of many minute crystallites increasing from pin point size to 1 mm or so in diameter with the distance from the tip of the rod. A few crystals then predominated. The underneath of the specimen which was in contact with the mould was less strongly polycrystalline than the free surface. For the fast rate of growth of 21 mm/min many crystal inserts were present along the rod which thus had some resemblance to the 'block' type of imperfection observed with bismuth (I, p. 566). For the medium rate of 5 mm/min the crystal form was not appreciably different from that of pure tin. For the slow rate of 0·5 mm/min the last few centimetres of the rod were, usually, a single crystal. In addition, as for pure tin, imperfections of lineage and fine lines on the surface of the specimen were observed. The latter were very marked.

For the 0·01% alloy, similar differences in the form of the first cm or so of the rods occurred, but in other respects the crystals were very similar to those for pure tin, single crystals being formed at both fast and slow rates. Fine lines were also observed, although these were not so prominent as for the 0·1% alloy.

It is concluded that 0·1% of cadmium prevents satisfactory single crystal growth, 0·01% does not, but surface imperfections are enhanced in both cases.

(ii) *Tin–Zinc*

Zinc (hexagonal structure, $c/a = 1{\cdot}86$) has a solid solubility in tin of approximately 2%. Two dilute binary alloys containing 0·1 and 0·01% of zinc were prepared from 99·997% pure zinc and 99·997% pure tin. The solid zinc tended at first to float on the liquid tin during the mixing but the alloy was subsequently shown spectrographically to be homogeneous. A negligible amount of zinc was lost by vaporization during the casting. The alloy rods were found to adhere strongly to the Pyrex casting tubes and also to show the characteristic non-specular reflection.

For the tin–zinc 0·1% alloy and for the fast rate of growth, crystals were produced, which were very reminiscent in form to bismuth crystals with block type imperfections and with inserts (I, p. 565). Crystals grown at the medium rate varied little from those of pure tin and, for the slow rate, only about half of the crystals obtained were single. As for the cadmium alloys the first centimetre or so of the rods was minutely polycrystalline and fine lines were prominent on the surfaces. The zinc was uniformly distributed along the specimens but more than 0·1% of zinc was found on some of the surfaces. For the 0·01% alloy little effect on the crystal form was observed; at both fast and slow rates single crystals were produced, but the concentration of zinc in the surfaces of the specimens was markedly greater than 0·01%.

It is concluded that the inclusion of 0·1% of zinc has a marked effect on the growth of single crystals, particularly for the fast rate of growth, but that 0·01% has little effect.

(iii) *Tin–Lead*

Lead (face-centred cubic structure) has a solid solubility in tin of 2% at the eutectic temperature. Two dilute binary alloys containing 0·1 and 0·01% of lead were prepared from 99·998% pure lead and 99·997% pure tin. The solid lead sank to the bottom of the liquid tin and appeared to be a little more difficult to dissolve than cadmium and zinc but, nevertheless, spectrographic examination showed the lead to be uniformly distributed.

For the tin–lead 0·1% alloy single crystal growth was prevented. For the fast rate inserts occurred and numerous crystallites with near orientations were observed, both at the beginnings and ends of the rods. For the slow rate, bicrystals were formed together with many crystallites at the beginning of the rods.

For the tin–lead 0·01% alloy, in contrast, only slight effects due to the lead were observed. Single crystals were formed at both the slow and fast rates and only one rod at the fast rate contained a number of inserts. Fine lines were not prominent on the surfaces of the crystals grown at the slow rate.

It is concluded that 0·1% of lead prevents single crystal growth but that 0·01% results only in minor changes in the form of the crystals.

(iv) *Tin–Silver*

Silver (face-centred cubic structure) has a limited solid-solubility in tin of approximately 0·1%. Two dilute binary alloys containing 0·1% and 0·01% of silver were prepared from fine silver wire (99·95% purity) and 99·997% pure tin. The mixing of the alloys was prolonged to ensure an even distribution of the silver and subsequent spectrographic examination showed that this was achieved. The cast rods as shown in fig. 1 were characteristic examples of the non-specular type reflection.

For the tin–silver 0·1% alloy the growth of single crystals was prevented. For the fast rate the rods contained inserts along their whole length, and for the slow rate the rods consisted of ragged strips of crystal. The surface imperfections were very prominent; the fine lines which occurred at the slow rate are shown in fig. 2 (Plate).

For the 0·01% alloy also the crystal growth was markedly affected. At the fast rate tiny crystallites were formed at places along the rods and at the slow rate half of the rods were bicrystals. The surface markings were similar to those of the 0·1% alloy. For both the silver alloys the spectrographic examination showed that the percentage of silver increased along specimens, there being about twice the quantity at the ends as at the beginnings.

It is concluded that silver prevents the growth of single crystals, 0·1% producing more imperfections than 0·01%.

(v) *Tin–Indium*

Indium (face-centred tetragonal structure, $c/a=1{\cdot}076$) is soluble to a considerable extent, probably 10%, in tin. Two dilute binary alloys containing 1% and 0·1% of indium were prepared from 99·9+% pure indium, kindly presented by Messrs. Mining and Chemical Products Ltd., and 99·997% pure tin. No difficulty was experienced in the preparation of the alloys and spectrographic tests showed the indium to be uniformly distributed. The indium spectra were found to be less sensitive to changes in concentration than the corresponding spectra of other metals. The indium alloys adhered to the Pyrex casting tubes and the surfaces of the rods showed prominent dendritic solidification patterns. Coating the inside of the tubes with a graphite suspension prevented adhesion of the metal to the glass surface.

For the tin–indium 1% alloy a marked difference in crystallization from that for pure tin was observed. For the fast rate of growth the rods were polycrystalline with inserts along their whole length, while for the slow rate a number of rods were single crystals and the others bicrystals. As usual, for the medium rate, the crystals were imperfect. On the surfaces fine lines, but much broader than those in any previous alloy, were observed, and for the fast rate dendritic surfaces were very noticeable.

For the 0·1% alloy the change in crystal form was less marked and even for the fast rate more than half of the crystals were single. Crystal inserts, and also crystals of near orientations, were present as imperfections. Single crystals, with a trace of lineage in one specimen only, were successfully grown at the slow rate.

It is concluded that, for a given percentage of alloying metal, indium has much less effect than the other metals previously considered; while for the fast rate 1% of indium prevents single crystal growth, 0·1% has little effect. Prominent surface markings occur for both percentages of the alloy.

(vi) *Tin–Antimony*

Antimony (rhombohedral structure) is soluble in tin to approximately 8% at 230° C. Two dilute binary alloys containing 1% and 0·1% of antimony were prepared from 99·92% pure antimony, given by the Department of Scientific and Industrial Research Pure Metals Committee, and 99·997% pure tin. The surfaces of the alloy rods did not show the marked non-specular reflection so characteristic of the other alloys. Etchants containing hydrochloric acid were found to be unsatisfactory and a dilute solution (1:20) of ferric chloride was used.

For the tin–antimony 1% alloy, the antimony had a considerable effect on the crystal form. For the fast rate the crystals consisted of numerous inserts, particularly on the underneath of the rod in contact with the mould. For the slow rate single crystals and bicrystals with prominent fine lines on the surfaces were obtained. For the 0·1% alloy the crystals differed little in form from those for pure tin. All the rods, except one at the fast rate which was a bicrystal and one at the slow rate which contained lineage, were single crystals. In fig. 3 (Plate) the dendritic surface, which is formed at the end of the crystals and which is more marked in the alloys than in the pure tin, is shown.

It is concluded that antimony, similarly to indium, has less effect as an alloying component on the growth of single crystals of tin than the other metals used.

(vii) *Summary*

In the table the forms of the crystals of the binary alloys are summarized. The results are given for both the fast and the slow rates of growth but not for the medium rate for which no marked difference from those of pure tin was observed. A rod was classified as a single crystal when the orientation of a continuous piece of it at least 15 cm long did not vary by more than 1°.

In the table brief descriptions consisting of one or two words are used but these should be understood as merely indicative of the main characteristics of the group.

Form of the Crystals of Tin Alloys

Alloying component	F	Fast rate 21 mm/min	Slow rate 0·5 mm/min
Cd 0·1%	0·1	Block imperfections and inserts	Nearly single crystals
Cd 0·01%	0·01	Single crystals (75%)	Single crystals (88%)
Zn 0·1%	0·05	Block imperfections and inserts	Nearly single crystals
Zn 0·01%	0·005	Single crystals (88%)	Single crystals (100%)
Pb 0·1%	0·05	Crystals with inserts	Bicrystals
Pb 0·01%	0·005	Single crystals (100%)	Single crystals (100%)
Ag 0·1%	1	Crystals with inserts	Strips of crystal
Ag 0·01%	0·1	Single crystals (50%) with inserts	Bicrystals and single crystals (50%)
In 1%	0·1	Crystals with inserts	Single and bicrystals
In 0·1%	0·01	Single crystals (88%)	Single crystals (88%)
Sb 1%	0·12	Crystals with inserts	Single and bicrystals
Sb 0·1%	0·012	Single crystals (75%)	Single crystals (88%)

The values given under F in column 2 of the table are the ratios of the percentage of alloying metal to the percentage of maximum solubility of the alloying metal.

In columns 3 and 4 percentages of single crystals are also given.

(viii) *Discussion*

From the table it will be seen that the alloys fall into two categories (*a*) those which contain alloying metal of approximately 0·1 of the solid-solubility limit and which, for the fast rate of growth, do not form single crystals but which, for the slow rate of growth, show defects in the single crystal growth, and (*b*) those which contain alloying metal of approximately 0·01 of the solid-solubility limit and which produce a high percentage of single crystals for both fast and slow rates of growth.

From these experiments, therefore, it appears that the effect of a metallic impurity on the growth of single crystals of tin depends on the solid solubility of the impurity. The effect depends also on the rate of growth, being greater for the fast rate than for the slow rate. For the medium rate (5 mm/min), however, there was no marked change from the imperfect crystals obtained previously with pure tin. It follows that under conditions where single crystals are not obtainable the effect of the added impurity cannot be observed—as perhaps occurred with the zinc crystals of various purities grown at fast rates by Schilling (1934).

For tin, an impurity of 0·01 of the solid-solubility limit has little or no appreciable effect on the single crystal growth, and higher proportions could be tolerated, particularly at the slow rate of growth. In other laboratories single crystals of various kinds of alloys, e.g. brasses, have been grown from the melt, but insufficient data exist for any useful comparison with the present results.

Silver was the least soluble of the alloying metals used and, in the tin–silver alloys, the concentration of silver was found to increase along the specimens. This is to be expected in view of the low solubility, since the silver would tend to move in front of the solidifying interface. For the other alloys the alloying metal was found, in general, to be evenly distributed along the specimens, but in some cases it was concentrated at the surface.

Fine lines, observed also on the surfaces of the pure tin crystals, were very prominent on the alloy crystals (fig. 2). For the slow rate of growth the fine lines were always visible (often, but not always, parallel to the basal plane) and became more marked with the higher proportions of alloying metal. Similar fine lines have frequently been observed for zinc (e.g. Rosbaud and Schmid 1925, Straumanis 1929, Smialowski 1937). For the indium and antimony alloys, in which 1% of the alloying metals was present, the fine lines appeared broadened and similar to those observed by Straumanis (1929). It is likely that the alloying metal tends to separate into layers, possibly eutectic layers, parallel to the direction of growth. Considerable evidence is available to show that these layers become finer and finally disappear in the pure unalloyed metal; compare, for example, the observations in the present work with those in I, and also the observations of Smialowski (1937) and of Jillson (1950) on zinc. The surfaces of the crystals grown at the fast rate of growth are characterized by the dendritic markings (fig. 3) which are more prominent than for the pure metal.

Some preliminary work on less pure lead and zinc was reported in I. For less pure, as compared with pure, zinc a slight but definite increase in imperfection was observed (I, p. 564). Between 99·99% and 99·998% lead no difference in crystal form was observed (I, p. 563) and the extra impurity may well be bismuth which is soluble in lead. For the 99·97% bismuth, the 0·004% iron which is insoluble may be partly responsible for the block type imperfection observed. However, the presence of this impurity cannot account for the lineage imperfection in the bismuth crystals grown at the slow rate, if the effect is similar to that in tin where the impurity has not been observed to increase the lineage imperfection. From the present experiments it would appear also that impurities in the 99·997% tin, considered up to now as basically 'pure' tin, have some slight effect on the crystallization and that higher percentages of single crystals could be prepared from even purer tin.

§4. THE CRYSTALLIZATION OF TIN IN CLOSED MOULDS

In I only open-end horizontal moulds were used. To study the effect of various constraints twenty-four rods of 99·997% pure tin were crystallized in the Pyrex moulds in which they were cast.

(i) *Method*

The tin was received in 1 kg ingots which had been cast in a marble mould. The metal was then cast, as for the alloys above, in three stages, into 60 cm lengths within 5 mm internal diameter Pyrex tubes, and these tubes were sealed off under vacuum. By careful heating and shaking each of these 60 cm lengths was divided into two, thus producing 24 specimens each one of which was sealed *in vacuo*. Twelve of these were crystallized with the furnace travelling horizontally. For each crystallizing run four enclosed specimens were placed together at the middle of the furnace tube. A thermocouple was fixed within the evacuated furnace tube at their midpoint and the temperature was recorded on a Kent potentiometric recorder. The furnace tube was evacuated in order to obtain the same conditions as those in I. A 1 cm space at the beginning of each mould allowed for expansion of the metal. From the results of the previous work on tin it was decided to use three rates of growth of 21, 5 and 0·5 mm/min. The furnace was 30 cm long and provided a medium temperature gradient of the order of 20° c/cm.

The other twelve specimens were used with the furnace travelling vertically. The furnace trolley was originally designed to be used, if required, in either the horizontal or vertical position. For the vertical arrangement (Goss 1952) the rails were aligned vertically with a plumb-line and a counterbalance for the weight of the furnace and trolley was added. For each crystallization run the furnace was moved upwards from the bottom of the furnace tube and passed completely over the moulds. The tin then commenced to crystallize from the bottom of the moulds and not from the unstable position at the top of the less dense liquid metal. The rates of growth and furnace conditions were identical with those used for the horizontal arrangement above.

Some preliminary experiments, however, with the vertical arrangement were unsuccessful. With too small a space between the metal and the bottom of the Pyrex mould the expanding metal splintered the mould and with too large a space the liquid metal fell and disturbed the crystallizing metal below. A distance of 10 mm between the end of the metal and the bottom of the mould was found to be satisfactory, and using this, no failure of the Pyrex moulds (with the exception of one mould which cracked) nor discontinuity in the crystals occurred.

The crystallized specimens were given a preliminary examination whilst in their moulds. The glass was then cut and broken carefully away. Defects due to this treatment were neglected in observations on the form of the crystals. Specimens were etched, examined and the orientations measured with the optical goniometer.

(ii) *Results*

The crystals grown in the closed moulds, for both the horizontal and vertical methods, were single for rates of growth of 21 and 0·5 mm/min but not for 5 mm/min. This is the same as for pure tin (I, p. 563).

For the fast rate, apart from the first few millimetres, each rod consisted of a single crystal with, in the horizontal method, a number of small inserts in the rod and, in the vertical method, traces of crystals of near orientations in the next centimetre. The surface markings associated with fast rates of growth were present and, for the vertical method, conical shrinkage cavities appeared in the top of the specimens.

For the medium rate, a variety of crystal forms was produced, varying from nearly single to several crystals separated by long irregular boundaries. Lineage imperfection was usually present.

For the slow rate, all the specimens except for the first few centimetres were single crystals. Occasional lineage was present in crystals grown by the horizontal method and, for those grown by the vertical method, slight variations in orientation of 1°–2° occurred in one or two specimens. Fine lines were observed on the surfaces of all the crystals.

The orientations of the crystals made in closed moulds did not differ from those previously observed for crystals grown in open moulds.

(iii) *Conclusion*

From these results it may be concluded that the restriction of the specimen in a glass mould and the weight of superincumbent metal at solidification have no marked effect on the crystallization of tin. No comparable results for other metals are available but the known experimental evidence, e.g. the many different methods which have been successfully used for single crystal growth, would appear to be in agreement. However, the results do not preclude the possibility that the vertical method tends in the case of the cubic metals to produce a preferred orientation. In this connection the results for lead (I, p. 567) may be compared with those of Tsuboi (1929) who used the vertical method.

ACKNOWLEDGMENTS

The author wishes to thank Professor A. M. Taylor and the University authorities for facilities provided, the Ministry of Education for a Further Education and Training Grant, and Mr. S. Weintroub for his encouragement and advice.

REFERENCES

Goss, A. J., 1952, *Thesis*, University of London.
Goss, A. J., and Weintroub, S., 1952, *Proc. Phys. Soc.* B, **65**, 561.
Jillson, D. C., 1950, *Trans. Amer. Inst. Min. Metall. Engrs.*, **188**, 1005.
Rosbaud, P., and Schmid, E., 1925, *Z. Phys.*, **32**, 197.
Schilling, H. K., 1934, *Physics*, **5**, 1.
Smialowski, M., 1937, *Z. Metallkunde*, **29**, 133.
Smith, D. M., 1948, *The Spectrographic Analysis of Tin and Tin–Lead Solders* (London: Tin Research Institute).
Straumanis, M., 1929, *Z. anorg. allgem. Chem.*, **180**, 1.
Tin Research Institute, 1949, *Equilibrium Data for Tin Alloys.*
Tsuboi, S., 1929, *Kyoto Coll. Sci. Mem.*, **12**, 223.

Variation of the Electric Strength of KCl and NaCl Crystals with Temperature

By J. H. CALDERWOOD* and R. COOPER
Electronics Department, University of Liverpool

MS. received 25th June 1952

ABSTRACT. The paper describes measurements of the electric strength of crystals of KCl and NaCl which were made using both d.c. and standard $\frac{1}{50}$ μsec unidirectional impulse voltages at temperatures between −74° c and +140° c. In this range of temperature, which included for both KCl and NaCl a transition from a region in which the electric strength increased with rise in temperature to one in which it decreased, no significant difference was found between values of electric strength obtained using d.c. and corresponding values obtained with impulses. This result contrasts with that of Suita for KCl. It supports qualitatively Fröhlich's theory of breakdown.

---

§ 1. INTRODUCTION

According to the theories of dielectric breakdown developed by Fröhlich (1939, 1947) there exists for a given dielectric a critical temperature T_c at which a change occurs in the mechanism of breakdown. At temperatures less than T_c the material is said to be in the 'crystalline' region of breakdown and its electric strength increases with rise in temperature. In contrast with this behaviour the electric strength decreases with rise in temperature at temperatures greater than T_c, and the material is then in the 'amorphous' or high temperature region.

In the case of alkali halide crystals qualitative confirmation of the theory has been provided by d.c. measurements made mainly by von Hippel and his collaborators (Buehl and von Hippel 1939, von Hippel and Lee 1941). However, recent experiments (Alger and von Hippel 1949), which may be summarized as follows, suggest that this confirmation was purely fortuitous.

Both with direct current and with sawtooth impulses of duration 10^{-3} sec, 10^{-4} sec and 10^{-6} sec Alger and von Hippel measured the electric strength of KBr crystals at temperatures between −160° c and +360° c. Although the results obtained with direct current differed appreciably from those obtained for KBr by previous workers (Buehl and von Hippel 1939, von Hippel and Maurer 1941, Austen and Whitehead 1940) qualitative confirmation of the theory was again obtained. But with impulse breakdown Alger and von Hippel discovered an apparently more complicated situation. The effect of varying the impulse duration is illustrated in fig. 1. Even with 10^{-3} sec impulses appreciable differences were observed between the breakdown strength of the crystals and corresponding d.c. values. Further reduction in impulse duration caused additional changes of electric strength, and with 10^{-6} sec impulses the breakdown strength increased linearly with rise in temperature and exceeded the corresponding d.c. values throughout the entire temperature range.

Alger and von Hippel accounted for their results by assuming that space charge distortion of the electric field occurred in the interior of the specimens.

---

* Now at the Royal Technical College, Glasgow, C.1.

The space charge was assumed to consist of two components, one due to field emission of electrons from the cathode and the other due to migration of positive ions in the crystal. The density of both space charge components was assumed to depend on the temperature and the duration of applied stress so that the resulting field distortion also depended on these factors. The former component of space charge was assumed to be primarily responsible for differences observed at low temperatures but, owing to the increased ionic mobility, the latter component was assumed predominant at high temperatures. It follows that only by using impulses of short duration can reliable measurements of the electric strength of ionic crystals be obtained and, in fig. 1, the curve corresponding to the 10^{-6} sec impulses was considered to represent most closely the variation with temperature of the electric strength of KBr crystals. In contrast with the d.c. results the curve provides no evidence of an 'amorphous' region of breakdown.

Alger and von Hippel made measuremens only on KBr crystals but it may be inferred that similar effects would occur in other alkali halide crystals. The measurements described below, which were made on crystals of KCl and NaCl with both d.c. and $\frac{1}{50}$ μsec impulses do not appear to support this view.

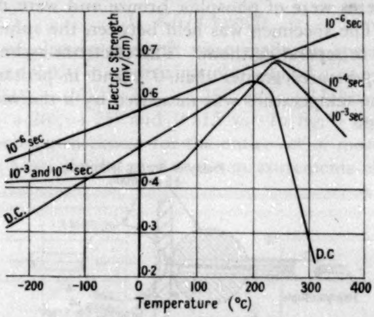


Fig. 1. Illustrating effect of impulse duration on electric strength of KBr (according to Alger and von Hippel).

§ 2. EXPERIMENTAL

2.1. *Preparation of Specimens*

The crystals from which the specimens were prepared were grown from molten salt of analar purity by a method based on that described by Kyropoulus (1926). In order to relieve any strain that may have been produced during its growth, each parent crystal was annealed by maintaining its temperature close to the melting point for approximately six hours and then allowing it to cool to room temperature over a period of about 12 hours. The parent crystals so obtained were cleaved along [100] planes to form rectangular plates of dimensions approximately 2 cm × 2 cm × 0·35 cm, and a test specimen was prepared from each plate in the following manner. In each plate, at the centre of one face, a spherical cavity about $\frac{1}{4}$ in. in diameter was produced by machining away the crystal. Thus at the centre of the plate its thickness was made a minimum and this was adjusted to the desired value, between about 0·005 in. and 0·03 in., by dissolving in water the crystal from both sides of the cavity. After this process some of the recessed specimens were subjected to an annealing process similar to that applied to the parent crystals, but since little difference was ultimately found between the behaviour of such specimens and that of specimens not subjected to this treatment, no distinction has been made between them in subsequent sections of the paper.

The surfaces of all specimens were highly polished, and each specimen was carefully examined to ensure that it contained no mechanical defects. The specimens were then stored in a desiccator until they were tested. The period of storage was usually less than 24 hours.

Before stressing a specimen its minimum thickness was measured to within 0·0001 in. by means of a dial-gauge micrometer, and then colloidal graphite electrodes were applied to the polished surface of the recess and its projection on the plane face of the specimen.

2.2. *Electrode Assembly*

Contact was made to the specimen by the electrode arrangement illustrated in fig. 2. The spheres were of phosphor bronze and were normally in contact with each other. The specimen was held between the spheres by light spring pressure developed in separating them. The electrode assembly was immersed in cable oil at temperatures greater than 0° c and in pentane at temperatures less than this. The temperature was measured by a thermocouple embedded in one of the spheres.

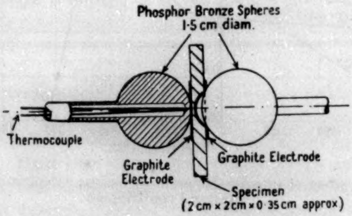


Fig. 2. Electrode arrangement.

The electrode arrangement just described is similar to that used by Austen and Whitehead (1940) and experience gained during the last few years has shown that with a suitable ambient it is satisfactory for the measurement of electric breakdown strength greatly in excess of any value yet reported for alkali halide crystals (Whitehead 1951).

2.3. *Measurement of Breakdown Voltage*

The d.c. source was of conventional design, the output voltage being measured by means of an electrostatic voltmeter and wire resistance potential divider. The voltage applied to the specimen was increased at the rate of approximately 1 kv/sec until breakdown occurred and was measured to an accuracy within about ± 3%.

A single stage resistance–capacitance impulse generator was used to generate the $\frac{1}{50}\,\mu$sec unidirectional impulses. It was charged from the d.c. set described above, and the peak impulse output voltage was determined from the efficiency of the generator and the d.c. charging voltage, measured as described previously. The former quantity was determined over the working voltage range by measuring the peak impulse voltage with a standard irradiated sphere gap.

To determine the impulse breakdown voltage of a specimen, the applied impulse voltage was increased by increments of 5% or less from about 70% of the estimated breakdown voltage until breakdown occurred. At each step a single impulse was applied to the specimen, and before applying the next impulse the polarity of the impulse generator was changed in order to minimize possible accumulation of ionic space charge caused by the successive impulses The breakdown voltage was thus known to lie between two values which were within 5% of each other. In the following sections of the paper the maximum of the two limits is given. The accuracy of peak impulse voltage measurement was probably within ±6%.

§ 3. RESULTS

3.1. *KCl Crystals*

Measurements of electric strength were made on specimens of thickness between 0·005 in. and 0·030 in. approximately, using d.c. at each of the following temperatures: −74°, +20°, +50°, +75° and +115° c, and unidirectional impulses at −74°, +20°, +75° and +115° c. In fig. 3 breakdown voltage is plotted against specimen thickness for the entire set of measurements obtained at −74° c. Figure 4 is a similar plot of the measurements obtained at +115° c.

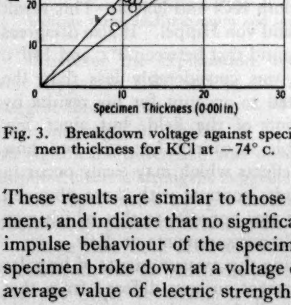


Fig. 3. Breakdown voltage against specimen thickness for KCl at −74° c.

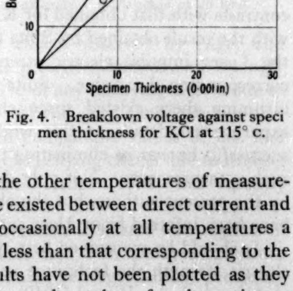


Fig. 4. Breakdown voltage against specimen thickness for KCl at 115° c.

These results are similar to those obtained at the other temperatures of measurement, and indicate that no significant difference existed between direct current and impulse behaviour of the specimens. Very occasionally at all temperatures a specimen broke down at a voltage considerably less than that corresponding to the average value of electric strength. Such results have not been plotted as they were considered due to faulty specimens. The total number of such specimens was not sufficient at any temperature to alter by more than one or two per cent the average value of electric strength.

The results plotted in figs. 3 and 4 also show that within limits of about ±15% the breakdown voltage varied directly as specimen thickness, which suggests that injurious discharges were absent in the ambient medium. From this evidence, coupled with the satisfactory appearance of the breakdown track in the region of minimum specimen thickness, it may be inferred that intrinsic electric strength was being measured (cf. Whitehead 1951, p. 11).

In fig. 5 electric strength is plotted against temperature. At each temperature the average value of electric strength, obtained from the total of both d.c. and impulse measurements is given, the scatter in the individual observations of electric strength also being indicated. The occasional low values mentioned previously were omitted in indicating the scatter.

The results of both d.c. and impulse measurements of electric strength on KCl exhibited qualitative agreement with the theory of Fröhlich, and indicate for this material a critical temperature of approximately 45° c.

3.2. *Sodium Chloride*

Very similar measurements to those just described were made on NaCl crystals, and the results obtained were of the same type. Measurements with d.c. were made at −74°, +20°, +50°, +75° and +140° c. Impulse measurements were made at −74°, +20°, +60° and +115° c. The results obtained are illustrated in fig. 6, and they provide no evidence that the impulse and d.c. electric strengths are significantly different in the temperature range of the experiments. In this diagram results obtained by Alger and von Hippel (1949) for NaCl are also given. It is not certain whether their curve is for average values or 'maximum' values of electric strength, but within reasonable limits agreement between the two sets of results is satisfactory.

As in the case of KCl, the results of the measurements are qualitatively in agreement with the theory of Fröhlich. The critical temperature was about 50° c.

§ 4. DISCUSSION

In the experiments described above, no significant differences were found in a temperature range that included T_c, between the direct current and the unidirectional $\frac{1}{50}\mu$sec impulse breakdown strengths of both KCl and NaCl. This result contrasts with that obtained for KBr by Alger and von Hippel. It also disagrees with the result obtained by Suita (1951) who found that between 0° c and 140° c the 1 μsec impulse electric strength of KCl was considerably less than the corresponding d.c. value. Suita also attempted to account for his results by assuming there existed space charge distortion of the field, but since few experimental details of this work are available it cannot be estimated how successful he was in eliminating the spurious effects which may easily occur in such measurements. Not only do Suita's results contrast with those obtained by the present authors, but the observed effect was opposite to that which may have been inferred from Alger and von Hippel's measurements on KBr.

It should be mentioned here that in obtaining the measurements of impulse electric strength the technique used by the present authors differed from that of Alger and von Hippel (1949). The latter applied to each specimen a single linearly rising voltage transient, so that breakdown occurred when the applied voltage was rapidly increasing. In the present experiments a number of impulses of increasing magnitude were successively applied until breakdown occurred on the crest of the wave, when the rate of change of voltage was small. Clearly when the duration of the impulse is comparable with the time taken for the discharge to develop, differences may be expected between the results of measurements made by the two methods. However, both theoretical (Fröhlich and O'Dwyer 1950) and experimental (Whitehead 1951, pp. 89–96) estimates of this time for solid dielectrics make it seem extremely unlikely that the different character of the two sets of results is due to this difference in technique.

A possible consequence of applying several impulses of the same polarity to obtain the breakdown voltage of a specimen is that ions may be caused to migrate by the successive impulses and thus lead to space charge field distortion and false results. The precaution taken of reversing the polarity of each successive impulse to minimize any net space charge distortion caused by the above effect has already been mentioned; with this procedure it does not seem likely that the net ionic migration caused by applying at most between 10 and 15 $\frac{1}{50}\mu$sec impulses would exceed that caused by a single sawtooth impulse of 10^{-3} sec duration. But with such impulses Alger and von Hippel found a considerable difference between the d.c. electric strength and the corresponding impulse electric strength of KBr.

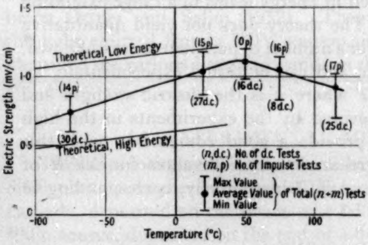


Fig. 5. Variation of electric strength of KCl with temperature.

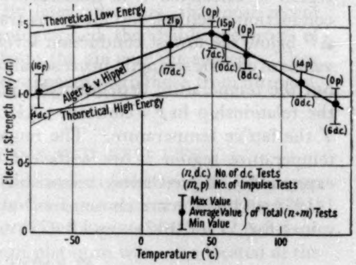


Fig. 6. Variation of electric strength of NaCl with temperature.

Numbers in brackets give numbers of d.c. or impulse (p) tests. Vertical and horizontal lines indicate range and maximum and minimum values of all tests. ● Average of all tests.

Whatever the explanation may be of the effects observed by Alger and von Hippel on KBr, and by Suita on KCl, the present measurements strongly suggest the absence between −74° c and +140° c of any falsifying effects due to ionic migration in crystals of KCl and NaCl. At the same time, the fact that no significant difference was observed between comparable values of d.c. and impulse electric strengths suggests the absence of falsifying effects due to heating of the specimens by pre-breakdown current. Keller (1951) has recently obtained evidence of this in glass.

Since with both KCl and NaCl the measurements indicated two regions of breakdown, one in which the electric strength increased with rise in temperature and the other in which it decreased, they must be regarded as providing qualitative confirmation of Fröhlich's theory.

Quantitative agreement with theory is however not very good. In fig. 5 values of electric strength for KCl in the crystalline region of breakdown are plotted according to the high electron energy criterion of breakdown, which was proposed by Fröhlich, and also according to the low electron energy criterion favoured by von Hippel and his collaborators. The latter curve is based on the formula given by Callen (1949). In calculating the values plotted on both curves the free space electron mass was used. This leads to values of electric strength equal to $1/\sqrt{2}$ of those which usually appear in the literature for the high energy criterion. Apparently the factor $\sqrt{2}$ was first introduced by Fröhlich who was under the impression that the electric strength of alkali halide crystals was anisotropic. There seems to be little justification for this assumption, and in the present calculations it has not been made. In fig. 6 similar theoretical curves are given for NaCl.

The measured values of electric strength for both KCl and NaCl do not decisively support either set of theoretical values. At 20° C they agree fairly well with values calculated on the basis of the low energy criterion of breakdown, but at $-74°$ C agreement with the high energy values is good.

The amorphous region of breakdown is distinguished by Fröhlich from the crystalline region by the assumption that in the former region mutual collisions between electrons are more frequent than collisions between conduction electrons and lattice vibrations. When the electric strength of a solid in the amorphous region is calculated it is necessary to consider the energy exchange between conduction electrons and electrons trapped in energy levels in a range extending ΔV below the lowest conduction level. The theory does not yield quantitative values of electric strength, because values for a number of parameters are unknown, but the temperature dependence is given, and can be tested experimentally by the relationship $\ln F = \text{constant} + \Delta V/2\boldsymbol{k}T$ where F is the electric strength and T the lattice temperature. The range covered by the experiments in the high temperature region is not sufficient to provide a good comparison with this expression. Nevertheless reasonable agreement exists if values for $\Delta V/\boldsymbol{k}$ of 1410° C and 830° C are chosen for NaCl and KCl respectively, corresponding to values for ΔV of 0·12 ev and 0·073 ev.

ACKNOWLEDGMENTS

The authors wish to acknowledge the assistance rendered by Mr. R. W. H. Stevenson, Department of Natural Philosophy, University of Aberdeen, who provided a crystal of KCl at the outset of the work.

REFERENCES

Alger, R. S., and von Hippel, A., 1949, *Phys. Rev.*, **76**, 127.

Austen, A. E. W., and Whitehead, S., 1940, *Proc. Roy. Soc.* A, **176**, 33.

Buehl, R. C., and von Hippel, A., 1939, *Phys. Rev.*, **56**, 941.

Callen, H. B., 1949, *Phys. Rev.*, **76**, 1394.

Fröhlich, H., 1939, *Rep. Prog. Phys.*, **6**, 411 (London : Physical Society); 1947, *Proc. Roy. Soc.* A, **188**, 521.

Fröhlich, H., and O'Dwyer, J., 1950, *Proc. Phys. Soc.* A, **63**, 81.

von Hippel, A., and Lee, G. M., 1941, *Phys. Rev.*, **59**, 824.

von Hippel, A., and Maurer, F. J., 1941, *Phys. Rev.*, **59**, 820.

Keller, K. J., 1951, *Physica*, **17**, 511.

Kyropoulus, S., 1926, *Z. anorg. Chem.*, **154**, 308.

Suita, T., 1951, *Tech. Rep. Osaka Univ.*, **1**, 51.

Whitehead, S., 1951, *Dielectric Breakdown of Solids* (Oxford : Clarendon Press).

The Luminescence of Air, Glass and Quartz under α-Particle Irradiation

By J. B. BIRKS and J. W. KING*
Department of Physics, Rhodes University, Grahamstown, South Africa

MS. received 29th August 1952

ABSTRACT. The luminosity of a ^{210}Po source has been detected using a photomultiplier technique. It is also found that photons are produced in air, glass and quartz when these materials are irradiated by α-particles. It is shown that the ' scintillations ' observed by Richards and Cole probably originated in these materials, and not in the various thin films to which they were ascribed.

---

§ 1. INTRODUCTION

It is well known that ^{210}Po and other α-particle sources are feebly self-luminous. The spectrum of this luminosity was first examined with a quartz spectrograph by Sir William and Lady Huggins (1903, 1905 and 1906) and found to coincide with that of nitrogen. They concluded that the effect was due to α-particle excitation of nitrogen occluded in the source, and in the air irradiated by the source. This has been confirmed by later observers, the most recent being Ortner and Salim (1952). They photographed the visible spectrum of the light emitted from a ^{210}Po source, and from the air near the source. They identified the various lines and bands in the two spectra with those from nitrogen, and also oxygen.

§ 2. EXPERIMENTAL METHOD AND RESULTS

This luminosity and similar effects have been investigated using an EMI type 5060 photomultiplier tube, which has a 1 cm diameter semi-transparent photocathode, deposited on the inside of a 1·5 mm thick flat glass window. A 2·4 mc ^{210}Po source, deposited on the end of a 0·9 mm platinum wire, was located at the end of a brass collimating tube, 15 mm long and 1·5 mm internal diameter, giving a narrow emergent beam of about 4×10^6 α-particles per minute. The source was mounted in air directly above the photo-cathode, on a micrometer screw, so that the distance from the tube could be varied. The output from the photomultiplier was fed through a cathode follower, linear amplifier and discriminator to a fast scaling unit. Throughout the experiments all surfaces were kept clean and free from materials not specifically mentioned.

With the source in position, a large number of small amplitude pulses, greatly exceeding the noise count, was observed. The integral pulse-height distribution curves gave no indication of a plateau, and comparative measurements were therefore made of the counting rate at constant gain and a fixed discriminator bias. The number of counts observed at this bias was approximately 0·18 of the total number of counts, extrapolated to zero bias.

Figure 1 shows the observations on the variation of counting rate with the distance of the source from the photomultiplier. This distance is expressed in

---

[*] Now at Emmanuel College, Cambridge.

millimetres of standard air equivalent (15°c, 76 cm Hg), the range of the ^{210}Po α-particles being 38·2 mm. The noise count of 370 counts/min has been subtracted throughout. The gross count curve ABC is attributed to counts arising from photons produced A at the source, B in the air between the source and the photomultiplier, and C in the glass of the photomultiplier.

The counts A were eliminated by covering the source with an aluminium foil of 1·3 mm equivalent air thickness, placed 0·1 mm in front of the source. Curve BC was then obtained, the distance being corrected for the equivalent air thickness of the foil. This curve shows two separate effects, one of which occurs only when the source is less than about 38 mm from the tube. Curve A gives the difference between curves ABC and BC, and represents the counts due to photons produced at the source.

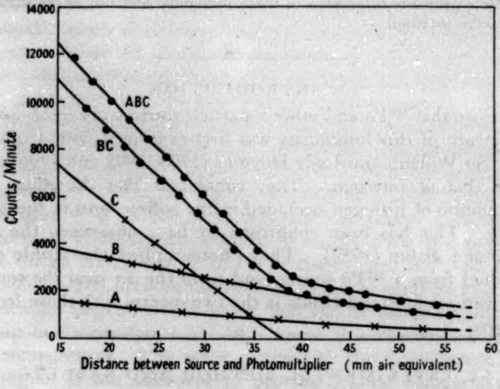


Fig. 1.
A. Counts due to photons originating at the source.
B. Counts due to photons originating in the air.
C. Counts due to photons originating in the glass.

The counts B, due to photons produced in the air, were next eliminated (fig. 2) by placing three aluminium foils, each of 1·3 mm air equivalent thickness, directly on the photomultiplier; curve II was obtained. When the distance of the source from the photomultiplier was corrected by the addition of the equivalent air thickness of the foils, curve I was obtained. It is seen that the photons cease to be detected when the source distance is greater than 38·2 mm air equivalent, i.e. greater than the α-particle range, and the number of photons detected increases with the residual range of the α-particles in the glass. Curve III was obtained when four more similar foils were placed on the photomultiplier. Curve III is parallel to curve II and coincides with it if the distance is corrected by the addition of the equivalent air thickness of the four extra foils. It is concluded that the photons are produced in the glass, and not in the aluminium foils, which have no effect on the number of counts observed. Curve C of fig. 1 shows the number of counts due to photons produced in the glass. Curve B, obtained by subtracting C from BC, represents the counts due to photons produced in the air.

These experiments were repeated with a fused quartz slab 1·45 mm thick placed on the photomultiplier. Photons produced in the quartz were detected, and separated from those produced at the source and in air, by introducing aluminium foils above the quartz. When an aluminium foil was placed between the quartz and the photomultiplier, the effect was completely eliminated. The results were similar to those obtained with the glass but the number of counts due to photons produced in the quartz was a factor of two less, and those due to photons produced in the air and at the source a factor of three less. This is attributed to the reduced solid angle of detection, and possible optical absorption in the quartz.

§ 3. DISCUSSION

It is concluded that the α-particles are responsible for the production of photons at the source, in air and also in glass and quartz. The effect cannot be ascribed to the weak γ-radiation from ^{210}Po, as the photons from the glass cease when the α-particles are unable to penetrate into it.

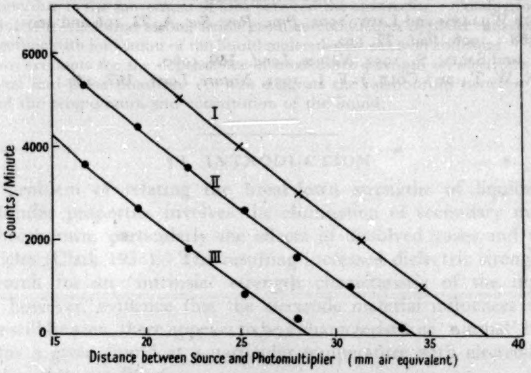


Fig. 2.
I. Counts due to photons originating in the glass.
II. With three aluminium foils between source and glass.
III. With seven aluminium foils between source and glass.

A rough estimate, made by extrapolating curve C of fig. 1 to zero source distance, indicates that the total number of counts due to photons that would be produced in the glass by 5·3 MeV α-particles is only 0·02 of the number of incident α-particles. Assuming a mean photo-cathode conversion efficiency of about 2%, this indicates that only about one photon per α-particle reaches the photo-cathode. The total number produced may, however, be several times higher, since the glass is opaque at wavelengths below 320 mμ. The results with the quartz indicate that it is reasonably transparent to the photons, and it would be desirable to repeat the measurements using a photomultiplier with a quartz envelope.

Richards and Cole (1951) have described experiments in which they observed 'scintillations' when thin films of many different materials were irradiated by α-particles from a strong ^{210}Po source. The 'scintillations' were only observed when the thickness of the film was less than the range of the α-particles in the material. In the majority of cases, the films were deposited on a 1 mm quartz slide, which was reversed to take noise readings. Richards and Cole referred neither to the self-luminosity of the source nor to the photons produced by α-particles in air, quartz or the glass of the photomultiplier envelope. It appears from their report that in each case the residual range of the α-particles, after emerging from the film, was in one or more of these three materials. It is evident that some, and probably all, of the 'scintillations' ascribed to the thin films originated in the quartz, glass or air. This would explain their statement that "the efficiencies of all these films are approximately the same", since the efficiency in question would, in the majority of cases, be that of the quartz slide below the film.

ACKNOWLEDGMENT

We are indebted to the South African Council for Scientific and Industrial Research for a grant in support of this work.

REFERENCES

Huggins, Sir William and Lady, 1903, *Proc. Roy. Soc.* A, **72**, 196 and 409 ; 1905, *Ibid.*, **76**, 488 ; 1906, *Ibid.*, **77**, 130.

Ortner, G., and Salim, S., 1952, *Nature, Lond.*, **169**, 1060.

Richards, E. W. T., and Cole, J. F. I., 1951, *Nature, Lond.*, **167**, 286.

Electrical Conduction and Breakdown in Liquid Dielectrics

By D. W. GOODWIN and K. A. MACFADYEN
Department of Physics, University of Birmingham

MS. received 23rd July 1952, *and in amended form 2nd October* 1952

ABSTRACT. Measurements of the breakdown strength of purified organic liquids have been made using rectangular voltage pulses of duration between 0·2 and 20 microseconds. The effects of cathode metal, γ-radiation and of changes in pulse duration, electrode separation, temperature, constitution of the liquid, etc. lead to a theory of breakdown according to which an instability arises as a result of electron emission from the cathode and field distortion due to the movement of ions between the electrodes. Conduction measurements support the view that strong fields produce conduction by field emission from the cathode, together with ionization of the liquid molecules by electron collisions. This theory of breakdown accounts for the dependence of breakdown strength on electrode separation, cathode metal and pulse duration. It also suggests the relationship between breakdown strength and the temperature and constitution of the liquid.

§ 1. INTRODUCTION

The problem of relating the breakdown strengths of liquids to their molecular properties involves the elimination of secondary mechanisms of breakdown, particularly the effects of dissolved gases and suspended solid particles (Clark 1933). The resulting increased dielectric strength has led to the search for an 'intrinsic' strength characteristic of the liquid only. There is, however, evidence that the electrode material influences the result, though, as will be seen, there appears to be a characteristic or 'normal' breakdown strength for a given liquid at a particular temperature with electrodes having a standard surface condition.

One of the chief difficulties in previous experimental determinations has been the very large spread in breakdown strengths even when measured by a single observer with the same apparatus. Nikuradse (1931) and Inge and Walther (1935) showed that the employment of impulses instead of continuously applied voltages reduced the large spread. Macfadyen and Edwards (1949), using rectangular voltage pulses, found an approximately constant and repeatable breakdown strength when the pulse duration exceeded a few microseconds, together with a considerably enhanced strength as the pulse duration was reduced. Thus the employment of pulses not only helps to overcome experimental difficulties inherent in the nature of liquids, but it also reveals time effects which connect breakdown in liquids with the movements of ions (Attwood and Bixby 1943).

The experiments to be described in § 3 employ the technique of Macfadyen and Edwards with certain improvements and extra precautions.

§ 2. MEASUREMENT TECHNIQUES

2.1. *Breakdown Strength Determinations*

Rectangular voltage pulses are applied to a test cell contained in a distillation apparatus similar to that described by Edwards (1951). The main features of the apparatus are (i) a completely enclosed distillation and filtering system, (ii) the availability of fresh electrode surfaces and a fresh sample of liquid for each determination, (iii) provision for continuous microscopic observation of the gap.

High voltage pulses are produced either from a hard-valve pulse generator with a voltage variable up to 11 kv and a pulse duration variable between 0·2 and 20 microseconds, or from a transmission-line generator with a voltage variable up to 35 kv and a fixed pulse duration of 11 microseconds. The latter generator consists of a high impedance (1900 ohms) coaxial delay cable charged to a high voltage and discharged through a resistance.

The electrode separation is measured by direct observation with a microscope having a micrometer eyepiece, the electrodes being illuminated from behind with collimated light. The electrodes ($\frac{3}{8}$ in. diameter spheres) are polished with a soft rotating mop using fine jeweller's rouge. They are heated in *n*-hexane to remove any grease from their surfaces, and after careful cleaning to remove dust particles are fitted into the electrode assembly and placed in the spark vessel.

2.2. *Reproducibility of Measurements*

In addition to the employment of voltage pulses and the careful removal of dust particles, further precautions are needed to reduce the scatter amongst comparable results. The preparation of the electrode surface greatly affects the breakdown strength. The use of any polishing agent coarser than the finest jeweller's rouge leads to a spread in results. Electrolytic polishing is satisfactory. Also, unless a very thorough de-greasing is given after polishing, the performance of repeated tests with different areas of the same electrodes yields values of breakdown strength which fall progressively by as much as 25% until a roughly constant value is attained. By using very fine rouge for polishing, degreasing the electrodes, carefully filtering the liquid and continuously examining the gap to remove dust particles, the scatter amongst comparable results has been reduced to less than 5%.

§ 3. EXPERIMENTAL INVESTIGATION OF BREAKDOWN

3.1. *The Breakdown Time Lag*

Previous investigators (Macfadyen and Edwards 1949) have shown that the breakdown strength of a pure liquid is independent of the time of application of the field above a certain time τ_0 (a few microseconds). Below τ_0 the breakdown strength rises rapidly. In other words, between the instant of applying the field and the instant of breakdown there is a time delay τ which equals τ_0 when the applied field is reduced to the minimum value $F_0{}^*$ to produce breakdown. This effect is illustrated in each curve of fig. 1, which shows a set of measurements on *n*-hexane with phosphor bronze electrodes, using the hard-valve pulse generator. When the pulse duration was set at a value less than τ_0 and the voltage was just sufficient to cause breakdown, the breakdown always occurred at the end of the pulse. When the pulse duration was greater than τ_0 breakdown occurred at a time τ_0. No time lags longer than τ_0 have been observed except when initiated by dust particles at field strengths just less than $F_0{}^*$.

For pulse durations less than τ_0 each of the curves in fig. 1 can be represented by the empirical formula

$$F^* = C_1/(\tau - \tau_\infty)^c \qquad \ldots\ldots(1)$$

where C_1, c and τ_∞ are constants for the curve in question. If the equation remained valid over the whole range of τ the breakdown strength would be infinite for a pulse of duration equal to τ_∞. The three constants are determined from the observations by assuming tentative values of τ_∞ and drawing a series of logarithmic plots, the value which gives a straight line is accepted as τ_∞, and τ_0 can then be calculated. The values of these constants for various gap lengths using *n*-hexane are given in table 1. Figure 1 shows that as the gap length is increased the minimum breakdown strength F_0^* decreases at first and then becomes constant, whilst the limiting breakdown time τ_0 increases progressively. This gives rise to a region where F_0^* is independent of gap length and pulse duration ; this value of F_0^* will be called the *normal breakdown strength* F_n^*.

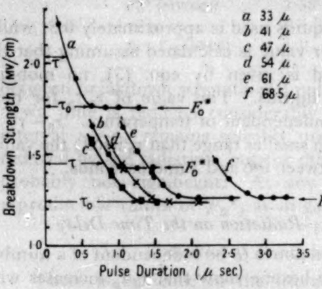


Fig. 1. **Variation of breakdown strength of *n*-hexane with pulse duration and gap length.**

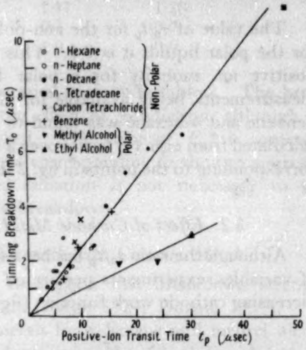


Fig. 2. **Relationship between the limiting breakdown time τ_0 and the positive-ion transit time t_p for various liquids. Phosphor-bronze electrodes were used.**

Two possible causes of time delay have been suggested: (i) the time taken for free electrons to receive sufficient energy to ionize the liquid molecules, (ii) the time taken for a low resistance path to be formed between the electrodes due to the movement of electrons and ions.

Only the second explanation predicts that the breakdown time will be a function of the gap length l and the positive ion mobility k_p. In table 1 the breakdown time τ_0 in *n*-hexane is compared with the positive-ion transit time t_p calculated from

$$t_p = l/k_p F_0^*. \qquad \ldots\ldots(2)$$

The value of mobility k_p (8·27 cm/sec per volt/cm) has been taken from the measurements of Adamczewski (1937). The ratio τ_0/t_p is approximately independent of the gap length, τ_0 being about one-third of t_p.

Determinations of this ratio have been carried out for a number of liquids of differing viscosities and hence different mobilities. The results are shown in fig. 2. The breakdown time τ_0 was varied for most of the liquids by using several different gap lengths and, in the case of benzene, different temperatures. The transit times were again calculated either from published experimental values of mobility (van der Bijl 1913) or by employing Adamczewski's formula relating mobility to viscosity:

$$k_p = C_2 \eta^{-3/2} \qquad \ldots\ldots(3)$$

where C_2 is a constant.

Table 1. Variation of the Time Lag in *n*-Hexane with Gap Length

Gap length l (μ)	33	41	47	54	61	68·5
τ_0 (μsec)	0·6	1·1	1·2	1·6	1·9	2·8
τ_∞ (μsec)	0	0·1	0·1	0·1	0·5	1·5
F_0^* (MV/cm)	1·76	1·43	1·24	1·24	1·26	1·26
C_1 (MV/cm)	2·7	2·6	2·5	2·6	2·6	2·0
c	0·3	0·3	0·3	0·4	0·4	0·2
t_p (μsec)*	2·26	3·46	4·6	5·25	5·85	6·55
τ_0/t_p	0·3	0·3	0·3	0·3	0·3	0·4
$\tau_0-\tau_\infty$ (μsec)	0·6	1·0	1·1	1·5	1·4	1·3

* Calculated from Adamczewski's value of mobility $8{\cdot}27\times10^{-4}$ cm/sec per volt/cm.

The value of τ_0/t_p for the non-polar liquids used is approximately 0·3, whilst for the polar liquids it is 0·1. This latter value is calculated assuming that the positive ion mobility for a polar liquid is given by eqn. (3), no mobility measurements being available for these liquids. The value of τ_0/t_p for both benzene and *n*-hexane was found to be independent of temperature. $\tau_0-\tau_\infty$ as calculated from eqn. (1) varies over a much smaller range than τ_0 itself; the values corresponding to the points in fig. 2 lie between 0·6 and 3 microseconds.

3.2. *Effect of Cathode Metal and γ-Radiation on the Time Delay*

Although the ratio τ_0/t_p has been shown above to be independent of a number of variables, experiments performed on *n*-hexane show that τ_0/t_p increases with increasing cathode work function (fig. 3 broken line). The electrodes used were not outgassed and values of the work functions were obtained by averaging figures quoted by Hughes and DuBridge (1932) for non-outgassed electrodes. It is probable then that the breakdown process in a liquid is affected by electron emission from the cathode as suggested by Young (1950) for liquids and highly compressed gases. As τ_0/t_p for both benzene and *n*-hexane is independent of temperature it is also probable that the electron emission is not of thermionic origin.

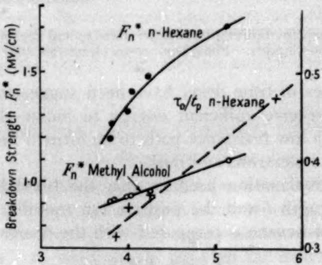


Fig. 3. Variation of F_n^* (full lines) and τ_0/t_p (broken line) with work function of cathode.

The irradiation of the cathode by intense ultra-violet radiation results in the breakdown strength of a gas being lowered (White 1936). An explanation in terms of field distortion by space charges has been given by Meek (1940). It is possible then that the reduction in τ_0 with decreasing work function is due to the quicker formation of space charge in the liquid.

Experiments on *n*-hexane were performed using either a 400 μc or a 5 mc radium source to irradiate the electrodes. The analysis of the relationship between breakdown strength and pulse duration is given in table 2. The breakdown time is reduced considerably by the effect of γ-radiation, although the interval $\tau_0 - \tau_\infty$ remains sensibly unchanged. The value of F_n^* is somewhat reduced by the radiation, but the difference corresponding to the two sources is probably not significant. At any rate, radiation is not necessary to give reproducible values of F_n^*, as in gaseous breakdown.

Table 2. Effect of γ-Radiation on Time Lag in *n*-Hexane

γ-radiation	None	400 μc	5 mc
c	0·3	0·3	0·3
τ_∞ (μsec)	0·7	0·3	0·2
τ_0 (μsec)	1·7	1·3	1·1
F_0^* (MV/cm)	1·38	1·17	1·20
τ_0/t_p	0·4	0·2	0·2
$\tau_0 - \tau_\infty$ (μsec)	1·0	1·0	0·9

3.3. *Variation of Breakdown Strength with Cathode Metal*

There is evidence in the literature to suggest that breakdown strengths depend on the cathode surface conditions (Sorge 1924, Dornte 1939). The variation of F_n^* with different cathode metals for *n*-hexane and methyl alcohol is shown in fig. 3 (full lines). The breakdown strengths of both liquids increase with increasing work function, suggesting that cathode emission is important in the breakdown process. A similar dependence on work function has been observed by Salvage (1951).

3.4. *Variation of Breakdown Strength with Gap Length*

The results illustrated in fig. 1 show that for *n*-hexane the lowest breakdown strength F_0^* is independent of gap length for values greater than 55 μ. For values less than this F_0^* increases with decreasing gap length. This is contrary to the results of Inge and Walther (1935) who, using millisecond impulses, showed that the breakdown strength was independent of gap length down to 2 μ. The present work shows that, for all the liquids investigated, F_0^* is inversely proportional to gap length l below a certain value of l. A typical set of results for chloroform is shown in fig. 4, with breakdown voltage $F_0^* l$ plotted against l. Below 60 μ the breakdown voltage is constant; above 100 μ the breakdown strength decreases slightly with gap length, a similar effect being observed with other liquids.

3.5. *Effect of Temperature on Breakdown Strength*

The normal breakdown strength F_n^* (§3.1) of several liquids has been examined and found to decrease with increasing temperature. The variation of F_n^* with temperature T for four paraffin hydrocarbons is shown in fig. 5, a plot of log F_n^* against $1/T$ giving a linear relationship represented by

$$F_n^* = C_3 \exp (C_4/T) \qquad \text{......(4)}$$

where C_3 and C_4 are constants. C_4 is independent of cathode metal whilst C_3 increases with increasing work function. C_3 also increases linearly with the homologue in the paraffin series, as suggested by Salvage on the basis of measurements at one temperature only. Similar measurements made with chlorine-substituted hydrocarbons (CH_2Cl_2, $CHCl_3$, CCl_4) show a progressive increase in C_3 with increasing chlorine content in the molecule.

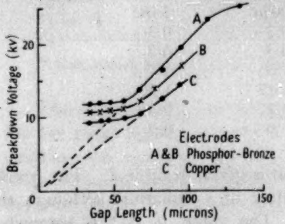


Fig. 4. Breakdown voltage of chloroform as a function of gap width ($\tau > \tau_0$).

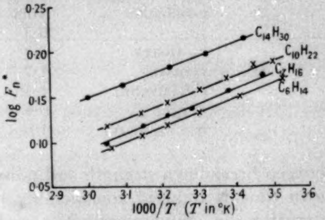


Fig. 5. Variation of normal breakdown strength of four paraffins with temperature (F_n^* in MV/cm).

3.6. *Discussion of Breakdown Measurements*

The dependence of normal breakdown strength F_n^* on cathode work function suggests that electron emission from the cathode is a contributory factor in breakdown. The absence of a statistical time lag confirms this view. Both F_n^* and its associated time lag τ_0 depend on the cathode work function in such a way as to suggest that cathode emission depends on field strength but not on temperature. The constancy of τ_0/t_p, in spite of variations in gap length, temperature, liquid, etc., indicates that breakdown is connected with the movement of ions. The different values of τ_0/t_p for non-polar and polar liquids may be due to either (*a*) wrong value for the positive ion mobility in a polar liquid calculated from eqn. (3) or (*b*) enhanced cathode emission due to the presence of a liquid of high dielectric constant. The variation of F_0^* with gap length (figs. 1 and 4) points to an electron multiplication process in liquid dielectrics.

§ 4. ELECTRICAL CONDUCTION IN LIQUID DIELECTRICS

4.1. *Theories of Conduction in Liquids*

The experimental evidence summarized in § 3.6 suggests that the important processes in breakdown are (*a*) electron emission from the cathode, (*b*) electron multiplication, (*c*) movement of positive ions in the field direction.

These processes must also be important in electrical conduction in highly purified liquids having no impurity ions present. Several entirely different theories of conduction in pure liquids have been published; these will now be reviewed briefly so that the evidence in favour of the proposed mechanism can be assessed.

Earlier investigators (Nikuradse 1932) obtained some evidence that ionization by collision occurred in liquids. Inge and Walther (1935) rejected this view on the grounds that the electronic free path in a liquid would be insufficient for an electron to acquire the energy (about 10 volts) needed for ionization. This objection was based on the idea that all collisions would be inelastic, but it stimulated a series of investigations designed to show that conduction in pure liquids was due entirely to cathode emission of electrons and their transport through the liquid (Baker and Boltz 1937, LePage and DuBridge 1940, Dornte 1940). Ruhle (1941), reverting to the collision-ionization theory, showed that the cathode emission giving rise to ionization by collision was field dependent. He assumed that this was due to ionization at the cathode–liquid interface, as previously suggested by Nikuradse (1932), but it may well be due to field emission made possible at relatively low field strengths by the influence of the liquid.

A theory not involving cathode emission was suggested by Plumley (1941) and investigated both by himself and Pao (1943). It is that the liquid acts like a weak electrolyte, and that an intense field encourages dissociation by reducing the binding energies within the molecule. The hypothetical dissociation process envisaged by Plumley

$$2C_7H_{16} \rightleftharpoons C_7H_{15}^- + C_7H_{17}^+$$

does not appear to be supported by other experimental evidence. Since an energy of 10·35 ev is needed to produce a similar positive ion ($C_7H_{16}^+$, Honig 1948), with no comparable compensation from the formation of the negative ion, the proportion of ionized molecules at room temperature would be only 10^{-179}. This is inadequate to account for the observed conduction currents at low field strengths.

Thus there is considerable evidence for electron emission from the cathode and no substantial evidence against ionization by collision. It is unlikely that electrons remain free in the liquid, but even if they become attached to molecules a possible explanation of collision-ionization is that, under the combined influence of thermal agitation and a strong field, the electrons occasionally break away and acquire enough energy to ionize a molecule.

None of the contending theories of conduction in liquid dielectrics has been extended to explain breakdown. It was therefore decided to investigate the conduction of *n*-hexane under conditions similar to those already used in breakdown measurements.

4.2. *Measurement of Conduction Currents*

Conduction measurements have been made using stabilized voltages variable up to 10 kv. It is believed that anomalies in current measurements observed by previous investigators are due to the attraction of small dust particles into the gap and also to ionic impurities. Considerable care has therefore been taken to remove such impurities.

Two highly polished plane parallel metal surfaces are supported in the liquid in such a way that their separation can be controlled from outside the containing cell. Electrical leakage is prevented by means of guard-rings of deposited gold in all positions where surface leakage could occur. The edges of the electrodes are rounded to prevent intense fields and continuous microscopic examination is provided. Liquid is distilled into the cell through a sintered glass filter and an electrolyser to remove ionic impurities, and is then returned to the distillation flask. The conduction currents are measured by passing them through high-stability resistors (10^6 to 10^{12} ohms) connected to an electrometer triode circuit. Currents down to 10^{-15} amp can be measured.

The liquid under investigation is placed in the distillation flask with a suitable drying agent and frozen with liquid air. The apparatus is then evacuated and, on heating, the liquid is allowed to distil around the system. A high voltage is applied to the electrolyser to remove ionic impurities. If any dust particles collect on the electrodes the liquid is emptied from the cell and fresh liquid passed through until no particles are visible. The presence of a particle during an experiment is usually detected by large fluctuations in the current. A test of the purity of the liquid is made at low field strengths (100 v/cm), for in this region the current density is roughly proportional to the applied field strength. A minimum conductivity of 10^{-18} ohm^{-1} cm^{-1} is usually obtained.

4.3. *Results of Conduction Measurements*

The variation of the conduction current in *n*-hexane with field strength for different gap lengths l using phosphor-bronze electrodes is shown in fig. 6. Each observation is the mean of five measurements, the spread being less than 10% if no dust particles are attracted into the gap. Previous investigators (Nikuradse 1932) reported that below 10 kv/cm the current density obeyed Ohm's law, and that this was followed by a region of saturation. However, this effect was not found provided a high field strength was applied to the liquid for several hours before the investigation, as found by Plumley (1941) and Pao (1943). The measurements are replotted in fig. 7 (full curves) as log i against l, where i is the current density calculated on the assumption of a uniform distribution over the electrode face. On the simplest theory of ionization by collision

$$i = i_c \, e^{\alpha l} \qquad \ldots\ldots(5)$$

where α is the ionization coefficient and i_c the density of electron current leaving the cathode; the logarithmic plot should therefore be a series of straight lines. At the lower field strengths the lines are fairly straight, though increasing curvature occurs with stronger fields. This effect has been attributed to recombination (Nikuradse 1932). The variation of α with field strength for different gap lengths (fig. 8) agrees with that found by Nikuradse, namely, α is proportional to E^2.

Extrapolation of the curves in fig. 7 to zero gap length (dotted portions) gives the emission density from the cathode i_c. In fig. 9 these values are replotted as log i_c/E_c^2 against $1/E_c$ and give a linear relationship. This variation of current with field strength is given by the Fowler–Nordheim (1928) relationship for field emission

$$i_c = aE_c^2 \exp(-b/E_c). \qquad \ldots\ldots(6)$$

Conduction measurements have been made on *n*-hexane using a different cathode material (fig. 7, broken lines). The cathode current is nearly an order of magnitude less with steel than with phosphor-bronze electrodes, but there is little change in the values of α. The variation of the cathode current with field strength agrees with eqn. (6). Values of the constants a and b are given in table 3. These constants are also compared with the Fowler–Nordheim constants for a metal–vacuum interface, assuming that the (current, field strength) relationship for a metal–liquid interface is as given by the Fowler–Nordheim eqn. (6). The constant a is 10^{-13} of the value for a metal–vacuum interface, whilst the constant b is 10^{-2} of the calculated value. Part of this large discrepancy is probably due to the lowering of the work function by the presence of a dielectric (Kalabuchow 1934) and to emission from points. It is also probable that a part is due to the small range over which observations have been made, the (current, field strength) relationship probably obeying a different law at high field strengths.

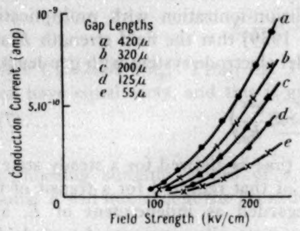


Fig. 6. Conduction current in *n*-hexane as a function of field strength.

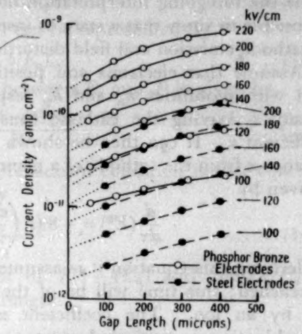


Fig. 7. Conduction in *n*-hexane plotted in accordance with eqn. (5), showing the effect of cathode metal.

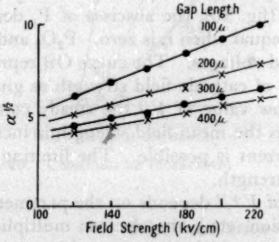


Fig. 8. Variation of apparent multiplication coefficient (α) of *n*-hexane.

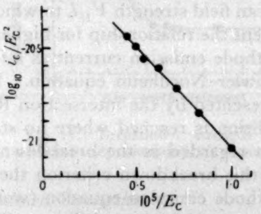


Fig. 9. Cathode emission I_c plotted in accordance with the Fowler–Nordheim equation. (E_c is in v/cm)

Table 3. Fowler–Nordheim Constants for *n*-Hexane from fig. 9 (ϕ in ev)

Cathode metal	In liquid Phos.-bronze	 Steel	In vacuo
a (amp/cm^2)	$2{\cdot}76\times10^{-20}$	$5{\cdot}1\times10^{-21}$	$1{\cdot}55\times10^{-6}/\phi$
b (volt/cm)	$2{\cdot}6\times10^{8}$	$2{\cdot}66\times10^{8}$	$6{\cdot}8\times10^{7}\phi^{3/2}$

Although there are differences between the observed and calculated values of the constants in eqn. (6), there is evidence to suggest that conduction currents at high field strengths are due to collision-ionization by electrons emitted from the cathode under the influence of the strong field. For the purpose of formulating a theory of breakdown it will be assumed that cathode emission is governed by the Fowler–Nordheim equation. Should subsequent research show that the (current, field strength) relationship is different, modifications can be made, but the basic principle will remain unchanged.

§ 5. THEORY OF BREAKDOWN

5.1. *The Mechanism of 'Normal' Breakdown*

If the foregoing interpretation of the conduction experiments is accepted it is possible to show that a state of instability can result from the combined effects of cathode emission and field distortion due to ionic movement.

Assume that electrons and positive ions move under the influence of the field with mobilities k_n and k_p respectively and that an electron current of density i_c leaving the cathode causes collision-ionization with multiplication coefficient α. It can then be shown (Loeb 1939) that the field strength E at a distance x from the cathode of a plane-parallel electrode system with gap length l is given by

$$\frac{d}{dx}(E^2) = -8\pi i_c \left(\frac{e^{\alpha l} - e^{\alpha x}}{k_p} - \frac{e^{\alpha x}}{k_n} \right). \qquad \ldots\ldots(7)$$

In deriving this equation it is assumed that time is allowed for a steady state to be reached; this time will be of the order of that required for a transit of the gap by an ion. The coefficient α is regarded as independent of E, and recombination is neglected since an approximate allowance can be made for this by using a reduced value of α (fig. 7).

Equation (7) is solved in § 5.2, where it is shown that for a constant voltage V between the electrodes the field strength E_c at the cathode rises linearly with i_c (eqn. (11)). This is represented by P_1Q_1 (fig. 10), the abscissa of P_1 denoting the mean field strength V_1/l, to which E_c is equal when i_c is zero. P_2Q_2 and P_3Q_3 represent the relationship for higher applied voltages. The curve OB represents the cathode emission current as a function of cathode field strength as given by the Fowler–Nordheim equation. For a low value of V/l the steady condition is represented by the intersection R_1, but as the mean field strength is increased a condition is reached where no stable current is possible. The limiting value V^*/l is regarded as the breakdown field strength.

In this breakdown criterion the value of V^*/l depends on the parameters of the cathode emission equation (work function, etc.), the electron multiplication coefficient and ionic mobilities. The way in which these quantities vary with temperature and constitution of the liquid will influence the breakdown properties of the system.

5.2. *The Condition for Breakdown*

Integration of eqn. (7) gives

$$E^2 = -8\pi i_c \left[\frac{e^{\alpha l} x}{k_p} - \left(\frac{1}{k_p} + \frac{1}{k_n} \right) \frac{e^{\alpha x}}{\alpha} \right] + E_0{}^2 \qquad \ldots\ldots(8)$$

where $E_0{}^2$ is a constant.

Insertion of the boundary condition

$$\int_0^l E\,dx = Fl, \qquad \ldots\ldots(9)$$

where F is the applied field strength, into eqn. (8) gives

$$E_0{}^2 = F^2 + 2\left[\frac{2\pi i_c e^{\alpha l} l}{k_p} - 8\pi i_c \left(\frac{1}{k_p} + \frac{1}{k_n} \right) \frac{e^{\alpha l}}{2\alpha^2 l} \right]. \qquad \ldots\ldots(10)$$

Near the breakdown conditions the values of α and l are such as to make the second term in the bracket much smaller than the first. For simplicity it will be neglected. By putting $x = 0$ and substituting eqn. (10) in (8) the cathode field strength E_c is given by†

$$E_c = F + 2\pi i_c\, e^{\alpha l} l/k_p F. \qquad \ldots\ldots(11)$$

For field strengths near breakdown eqn. (6) can be represented by

$$i_c = aE_c^2(1 - b/E_c). \qquad \ldots\ldots(12)$$

From eqns. (11) and (12) the relationship between E_c and F is given by

$$E_c - F = \frac{C_6 E_c^2 - bC_6 E_c}{F} \qquad \ldots\ldots(13)$$

where $C_6 = 2\pi a e^{\alpha l} l/k_p$. From fig. 10 the breakdown condition is that eqn. (13) shall have equal roots, and this is given by

$$F = \frac{2bC_6(\sqrt{C_6} - 1)}{1 - 4C_6}. \qquad \ldots\ldots(14)$$

By choosing suitable values of α, a plot of $F(1 - 4C_6)$ and $2bC_6(\sqrt{C_6} - 1)$ against F will intersect at the breakdown strength.

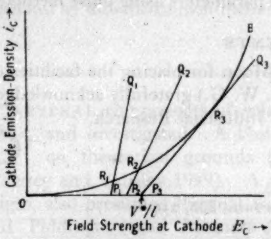

Cathode Emission-Density i_c →
B
Q_1
Q_2
Q_3
R_1
R_2
R_3
P_1
P_2
P_3
0
V^*/l
Field Strength at Cathode E_c →

Fig. 10. Condition for breakdown.

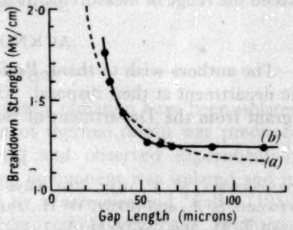


Fig. 11. Breakdown strength of *n*-hexane : (*a*) calculated, (*b*) observed.

5.3. *Calculation of Breakdown Strengths*

In fig. 11 a comparison is made between the breakdown strength of *n*-hexane as a function of gap length (*a*) as calculated and (*b*) as found experimentally. The variation of α with field strength for different gap lengths has been obtained from fig. 8. Had α been independent of gap length for constant field strength, then the breakdown voltage $F_n{}^*l$ would also have been independent of gap length. The effect of recombination is to make $F_n{}^*$ independent of gap length for values greater than about 70 μ. Although the agreement between theory and experiment is partly fortuitous, values of α and i_c having been extrapolated over a wide range of field strengths, it is significant that the general form of the curve is correctly predicted.

The Fowler–Nordheim constant *a* is inversely proportional to the cathode work function. Using values of *a* from table 3, the breakdown strength of *n*-hexane has been calculated for two different cathode metals. The observed and calculated breakdown strengths are compared below:

Cathode	phosphor bronze	steel
Calculated strength (MV/cm)	1·27	1·58
Observed strength (MV/cm)	1·24	1·48

---

† A similar equation has been developed by Meek (1940) for the case of a gas.

The calculations presuppose the attainment of a steady state of field distortion, which requires a time of the same order as the transit time of an ion. For breakdown with very short pulses a higher voltage must be applied in order to produce the necessary field distortion in the short time available. The rate of rise of breakdown strength with decreasing pulse duration will also depend on the variation of α with field strength in the high field region.

The variation of breakdown strength with temperature and molecular structure will probably depend on the variation of α with these parameters. It is significant that the admixture of chlorine-substituted hydrocarbons with paraffins, which may be expected to decrease α, has been shown (Ruhle 1941) to increase the breakdown strength.

5.4. *Conclusion*

A direct confirmation of the above theory would be to demonstrate experimentally the increase in cathode field strength which had been predicted. Work is at present in progress to produce this evidence. It is also intended to extend the range of measurements of α and other parameters using pulse methods.

ACKNOWLEDGMENTS

The authors wish to thank Professor P. B. Moon for placing the facilities of the department at their disposal. One of us (D. W. G.) gratefully acknowledges a grant from the Department of Scientific and Industrial Research.

REFERENCES

Adamczewski, M. J., 1937, *Ann. Phys., Paris,* **8,** 309.

Attwood, S. S., and Bixby, W. H., 1943, *J. Franklin Inst.*, **235,** 259.

Baker, E. H., and Boltz, H. A., 1937, *Phys. Rev.*, **51,** 275.

van der Bijl, H. J., 1913, *Deutsch. Phys. Gesell. Verh.*, **15,** 210.

Clark, F. M., 1933, *J. Franklin Inst.*, **216,** 429.

Dornte, R. W., 1940, *Ind. Eng. Chem.*, **32,** 1529.

Edwards, W. D., 1951, *Canad. J. Phys.*, **29,** 310.

Fowler, R. H., and Nordheim, L., 1928, *Proc. Roy. Soc.* A, **119,** 173.

Honig, R. E., 1948, *J. Chem. Phys.*, **16,** 105.

Hughes, A. L., and DuBridge, L. A., 1932, *Photoelectric Phenomena* (New York: McGraw-Hill).

Inge, L., and Walther, A., 1935, *Tech. Phys. U.S.S.R.*, **1,** 539.

Kalabuchow, N., 1934, *Z. Phys.*, **92,** 143.

LePage, W. R., and DuBridge, L. A., 1940, *Phys. Rev.*, **58,** 61.

Loeb, L. B, 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York: Wiley), p. 387.

Macfadyen, K. A., and Edwards, W. D., 1949, *Nature, Lond.*, **163,** 171.

Meek, J. M., 1940, *Proc. Phys. Soc.*, **52,** 547.

Nikuradse, A., 1931, *Arch. Elektrotech.*, **25,** 826; 1932, *Z. Phys.*, **77,** 945.

Pao, C. S., 1943, *Phys. Rev.*, **64,** 60.

Plumley, H. J., 1941, *Phys. Rev.*, **59,** 200.

Ruhle, F., 1941, *Arch. Elektrotech.*, **35,** 490; 1943, *Phys. Z.*, **44,** 89.

Salvage, B., 1951, *Proc. Instn. Elect. Engrs.*, Monograph II.

Sorge, J., 1924, *Arch. Elektrotech.*, **13,** 189.

White, H. J., 1936, *Phys. Rev.*, **49,** 507.

Young, D. R., 1950, *J. Appl. Phys.*, **21,** 222.

Thermal Theories of the High-Intensity Components of Solar Radio-Frequency Radiation

By J. H. PIDDINGTON

Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization, Australia

MS. received 20th May 1952, and in final form 21st October 1952

ABSTRACT. The high-intensity components of solar radiation are defined as those components whose emitting regions occasionally attain brightness temperatures of 10^{10} deg K or more. They have been referred to individually as enhanced radiation (or noise storms), outbursts and isolated bursts.

Theories of origin in terms of thermal processes are examined and none are found capable of explaining the observational data. It is concluded that all three components are generated by the ordered (non-thermal) motion of electrons.

---

§ 1. INTRODUCTION

Several components of solar radio-frequency radiation have been isolated and investigated. A basic component of thermal origin was predicted on theoretical grounds (Martyn 1946) and observed experimentally (Pawsey and Yabsley 1949). A slowly varying component was isolated and its origin, also presumed thermal, was discussed by Waidmeier and Muller (1950) and Piddington and Minnett (1951). These components have brightness temperatures ranging up to about 10^7 deg K.

Besides these thermal components there are others characterized by extremes of variability and intensity. As will be seen below, brightness temperatures of 10^{10} to 10^{11} deg K or more have been reached by each of these components which are referred to collectively, therefore, as the 'high-intensity components of solar radiation'. Individually they may be referred to (Pawsey 1950) as enhanced radiation (or alternatively, noise storms), outbursts and isolated bursts.

The origin of these components is not yet understood. Theories based on ordered (non-thermal) motion of electrons are generally favoured, but there are several which attempt to explain the observed radiation in terms of purely thermal processes.

In an ionized gas such as the solar corona, thermal radio-frequency radiation is emitted and absorbed by electrons suffering *random* accelerations by electric or magnetic fields. The intensity of emission cannot exceed that of a black body at the gas temperature. If the gas is in local thermodynamic equilibrium then its temperature is related in a well-known manner to the average kinetic energy of the gas particles. In the absence of local thermodynamic equilibrium, as might be anticipated in the solar corona, the concept of temperature may be retained. Its value is similarly related to a *suitably weighted* mean particle kinetic energy.

A significant difference arises if the accelerations of the charged particles are so *ordered* that the radiation emitted by each is in phase coherence. The intensity of emission is not then limited by (or even significantly related to) the temperature of the radiating particles. A striking example of radiation by ordered or non-thermal motion of electrons is that of a radio aerial emitting waves of very high intensity while remaining quite cold.

It is the purpose of the present paper to discuss the *thermal* theories of origin of the high-intensity components.

§ 2. THE OBSERVATIONAL DATA

Any satisfactory theory of origin of a component of solar radiation must show reasonable agreement with the observational data which is briefly reviewed here.

Most important, perhaps, is the maximum brightness temperature; in thermal theories this provided a lower limit of the temperature of the radiating gas. Values of flux density at the earth up to 10^{-18} watts $m^{-2}(c/s)^{-1}$ at 97 Mc/s have been recorded for enhanced radiation (Payne-Scott and Little 1951) and similar values for isolated bursts (Payne-Scott, private communication), while much higher values, up to 10^{-15} watts $m^{-2}(c/s)^{-1}$ at 60 Mc/s have been found for outbursts (Pawsey 1950). The size of the source of each of these components is known to be about one per cent or less, of the solar disc, giving brightness temperatures of 10^{12} deg K or more for enhanced radiation and isolated bursts and 10^{15} deg K for outbursts.

Another important characteristic is the maximum rate of change, particularly of decay, of the component. Enhanced radiation, which usually persists for periods of hours or days, appears to consist of bursts rising from time to time above a relatively steady base level. Bursts may have decay periods less than one second and the base level of some minutes (Wild 1951). Outbursts which are highly irregular last for some minutes and often show substantial changes in a few seconds. Since refraction seems negligible, as indicated below, major intensity variations probably occur at the source.

The observed spectrum of enhanced radiation is variable, sometimes being reasonably uniform over the frequency range (70–130 Mc/s) investigated by Wild (1951). On several occasions, however, it showed a maximum near 100 Mc/s with the intensity decreasing on either the high- or low-frequency side (or both) to half-value in 30 Mc/s or less. In the case of outbursts and bursts (either noise storm or isolated) the fall of intensity from maximum to zero may occur within 2 to 10 Mc/s either on the high- or the low-frequency side of the spectrum (Wild 1950 a, b, 1951).

Observing at 97 Mc/s Payne-Scott and Little found enhanced radiation arising from about 0·3–1·0 solar radii above large sunspots, usually spots which had reached their maximum size. They consider refraction effects unimportant and that the sources are correctly located. Outburst sources move radially with velocities of about 10^3 km sec^{-1} (Payne-Scott and Little 1952).

Enhanced radiation, observed by Payne-Scott and Little (1951), usually shows nearly circular polarization. The sense of polarization and polarity of the sunspot concerned are such as to indicate that the observed magneto-ionic component is that called the 'ordinary ray' (see § 4). Outbursts and bursts usually show random polarization (Payne-Scott 1949, Payne-Scott and Little 1952), but on occasions outbursts may show marked circular polarization during part of their lifetime.

§ 3. EARLY THEORIES

Kiepenheuer (1946) and Denisse (1947) have estimated the power radiated by electrons spiralling in the magnetic fields of sunspots. They have not, however, considered re-absorption effects and consequently have found values much higher than full black-body radiation for the assumed electron temperature. As the electron motion is random this result appears inadmissible.

Saha (1946) and Saha, Bannerjea and Guha (1947) have suggested that the ordinary magneto-ionic component of solar radiation might escape from levels below the photosphere where the temperature was very high. Piddington and Minnett (1951) have shown, however, that this is only so for a vanishingly small range of directions about that of the magnetic field and so may be neglected.

The thermal theory proposed by Waldmeier and Muller (1950) to account for the low-intensity slowly varying component appears to have been extended by them to include also enhanced radiation. This is obviously inadmissible because the brightness temperatures associated with this component may greatly exceed the maximum electron temperatures postulated by the theory. It seems probable as Piddington and Minnett (1951) have shown that there are two different components involved.

§ 4. RYLE'S THEORY

Ryle (1948, 1949 a, b) has suggested that *all* solar, as well as galactic, radio-frequency radiation has its origin in thermal or 'equilibrium' processes. This theory was recently revived (Ryle 1950) and will be discussed here in some detail.

(i) *The Theory*

An attempt is made to account for brightness temperatures up to 10^{10} deg K on the Sun and 10^{14} deg K on stars in terms of radiation by electrons having mean kinetic temperatures of similar magnitudes. According to the theory, the fast electrons are provided by electric fields induced by the movement of gases across magnetic fields. Such movement results from the non-uniform solar rotation and the solar general magnetic field and may be assisted by sunspot magnetic fields. Radiation results from collisions of the fast electrons with other particles.

Emission from an ionized gas is determined by the three quantities $x=4\pi N\boldsymbol{e}^2/\omega^2\boldsymbol{m}$, $y=H\boldsymbol{e}/\omega\boldsymbol{mc}$, $z=\nu/\omega$ where N, $\boldsymbol{e}$, $\boldsymbol{m}$ and ν are the electron density, charge, mass and collision frequency respectively, ω is the wave angular frequency, H the magnetic field strength and c.g.s. electrostatic units are used. The actual form of the relevant magneto-ionic equations depends also on the angle θ between the direction of propagation and the magnetic field.

Ryle considers that the longitudinal approximations should be used and that emission is mainly from regions immediately above the reflecting levels defined by $x=1-y$ (extraordinary ray), $x=1+y$ (ordinary ray) and $y=1$ (the gyro-frequency level, extraordinary ray).* Ryle considers that escape of radiation from the $y=1$ level past the (higher) $x=1-y$ level is not possible and proceeds to investigate emission from the other two regions.

---

* The use of the longitudinal approximations is valid for reasonably large values of θ except in the region where $x\sim1$ where the transverse approximations hold for all but vanishingly small values of θ and $x=1$ defines a level of reflection. Thus the ordinary ray from the (lower) $x=1+y$ level is blocked and the observed ordinary ray originates above the $x=1$ level (Piddington and Minnett 1951). Absorption above the $x=1$ and $x=1+y$ levels are of the same order (see, for example, Westfold 1951) so that the conclusions are not appreciably affected.

Expressions are found for the optical depths of the two radiating regions in terms of ν and the gradients of x and y. The emissivity is proportional to ν. Values were adopted of electron temperature and density of 10^{10} deg K and 10^8 cm^{-3} at the extraordinary ray level and 10^8 deg K and 10^{10} cm^{-3} at the ordinary ray level, and a value of the collision cross section of a proton for an electron of 10^{-17} cm^2. The assumed gradients of x and y are based on accepted electron density and sunspot magnetic field distributions. These data lead to values for ν of about 1 sec^{-1} and 500 sec^{-1} for the two levels, and resulting values of brightness temperatures of 6×10^8 deg K and 2×10^7 deg K at the extraordinary ray and ordinary ray levels respectively. Ryle considers that these estimates may be too low.

(ii) *Electron Collision Frequencies*

The principal objection to the above theory is that the electron collision frequencies have been grossly overestimated. This was pointed out briefly by Piddington and Minnett (1951*) and mentioned by Ryle at a conference on 'The Dynamics of Ionized Media' held in London during 1951. The use of revised estimates of collision frequencies leads to brightness temperatures of 3000 °K and 20000° K for the two regions dealt with above, instead of 6×10^8 deg K and 2×10^7 deg K as found from the theory.

Although the theory as originally formulated must be abandoned it is desirable to consider whether any possible distributions of electron temperature and density allow the theory to be reconciled with observations. For example, emissivity is increased by decreasing the gradients of x and y. However, it may be shown that the effective radiating region would have to be increased from about 10^3 km to about 10^9 km which is impossibly high. An overall increase in electron density is also ineffective since this is limited by reflection when $x \sim 1$. It is concluded that no reasonable values of the various parameters will provide brightness temperatures as high as 10^{10} deg K and that the theory must be rejected.

Further evidence against the theory is listed in the following section. This evidence is discussed in connection with radiation from the gyrofrequency level but much of it may be applied equally to the above theory.

Although not directly relevant to the present discussion it may be desirable to mention a theory of support of the solar atmosphere advanced by Ryle (1948) as a by-product of his thermal theory. The mechanism suggested is the reaction from radiation emitted downwards from the $y=1$ level. In estimating the intensity of such radiation the emission from individual electrons is summed, re-absorption and phase differences being neglected. The resulting intensity is far greater than black-body radiation at the adopted gas temperature. If the black-body value (the maximum permissible) is adopted, the supporting force is found to be negligible.

§ 5. RADIATION BY ELECTRONS ACCELERATED BY MAGNETIC FIELDS

The major difficulty met by Ryle's theory is that the emissivity of the radiating gas decreases with collision frequency. Such is not the case for radiation from the gyrofrequency level (where $y=1$) where collisions are not necessary to provide radiation. Giovanelli (1948) has proposed a theory of high-intensity solar radiation in terms of radiation from this level. He mentions electron energies of 10^6 ev which would be capable of providing brightness temperatures up to about 10^{10} deg K. However, as Ryle (1948) and, in more detail, Piddington and Minnett (1951) have shown, as long as Appleton's magneto-ionic equations hold such radiation is *always* prevented from leaving the solar atmosphere.

---

* Also read at the Zürich conference of Union Radio Scientifique Internationale, 1950.

There are two conditions for which these equations cease to hold. The first is for sufficiently low electron density (so that the spacing is an appreciable fraction of a wavelength), when the blocking layer ceases to exist and radiation may escape freely. This condition is unlikely ever to apply in the solar atmosphere as may be seen from the data discussed by Smerd (1950). The second condition is when the electron velocity is so high that the conditions of radiation are altered. The extreme case, when the kinetic energy of the electron is much greater than its rest mass energy, has been treated by Schwinger (1949) who shows that energy is radiated at frequencies much above the gyrofrequency*; such energy might be expected in general to escape from the solar atmosphere. This particular form of thermal emission was proposed by Alfvén and Herlofson (1950) to explain galactic radio emission as due to cosmic rays moving in magnetic fields. Ryle, at the London conference on 'Dynamics of Ionized Media', suggested that solar thermal radiation might escape in the same way; this is contrary to his earlier (Ryle 1948) conclusion.

It is also possible that radiation from less energetic electrons than those considered by Schwinger might escape from the gyrofrequency level because of Doppler shift effects. However, the components of radiation under consideration have brightness temperatures as high as 10^{12} deg K (enhanced radiation and isolated bursts) up to 10^{15} deg K (outbursts). The corresponding electron energies must be at least 10^8 to 10^{11} ev, values which would appear to qualify them for the Schwinger mechanism.

It would appear, therefore, that this mechanism offers the most likely explanation of high-intensity radiation in terms of thermal processes. A comparison is made below of the characteristics of radiation predicted by this theory and of the observed characteristics. For the purpose of completeness a brief discussion of the properties of radiation emitted by the Doppler shift process is included.

(i) *The Spectrum of Radiation*

The spectrum from a single electron radiating by the Schwinger mechanism has a broad maximum at a frequency say f_{m}. The intensity has only fallen by about 5% at frequencies of 30% above and below f_{m} and by a factor of 2 at frequencies of approximately $0{\cdot}05f_{\mathrm{m}}$ and $4f_{\mathrm{m}}$. The combination of spectra of a number of electrons having different values of f_{m} must show an even more gradual fall in power.

The observed decrease in intensity of enhanced radiation is often by a factor of 2 or more for frequencies $0{\cdot}7f_{\mathrm{m}}$ and $1{\cdot}3f_{\mathrm{m}}$. This corresponds to a rate of decrease between 10 and 60 times faster than that predicted by theory. It is possible that increasing absorption effects in the low-frequency range might account for a more rapid fall. If so, then the reverse should be true at the other end of the spectrum. It is concluded that the observed spectra are sometimes quite inconsistent with an origin by the Schwinger mechanism. In the case of bursts and outbursts the discrepancy is greater by an additional large factor (between 10 and 100 or more).

---

* The actual gyrofrequency of the fast electrons falls *below* that for slow electrons (given by $y=1$) because of the relativistic mass increase. Nevertheless, the frequency of maximum emission rises even more rapidly *above* that given by $y=1$.

The anticipated spectrum of radiation escaping by Doppler shift is less definite, depending on gradients of magnetic field and gas temperature. In the case of bursts and outbursts, however, it may be said that the observed spectra would indicate improbably high temperature gradients.

(ii) *The Rates of Fluctuation of Intensity*

Little may be said about the theoretical rates of increase of intensity of solar emission as these will depend on the unknown heating process. Rates of decrease of intensity may, however, be estimated from theoretical rates of cooling. At the high electron energies concerned collisions may be neglected and the particles assumed attached to the magnetic lines of force. The only appreciable loss of energy is due, therefore, to the radiating process itself.

Schwinger (1949) has shown that an energetic electron moving with angular velocity ω_0 in a circle radiates at a rate $P=2\mathbf{e}^2\omega_0{}^2\alpha^4/3\mathbf{c}$ where α is the ratio of the kinetic to the rest mass energy and is large. The frequency ω_m at which most energy is emitted is given by: $\omega_m=\frac{1}{2}\omega_0\alpha^3$ so that $P=8\mathbf{e}^2\omega_m{}^2/3\mathbf{c}\alpha^2$. The total energy of the electron is $\alpha\mathbf{mc}^2$ so that the time taken to lose all its energy at the above rate is given by $T=(\frac{3}{8}\mathbf{c}^3\mathbf{m}/\mathbf{e}^2)\alpha^3/\omega_m{}^2$. A reasonable value of ω_m for any of the components of radiation concerned is $2\pi\times10^8$ radians/sec so that $T\sim10^5\alpha^3$ second. Observed brightness temperatures correspond to values of α of at least 100 for enhanced radiation and isolated bursts and much higher for outbursts. This would correspond to decay times of thousands of years instead of a few minutes or seconds as observed.

Similar difficulties (although not quite so extreme) are met in trying to explain observed intensity decreases if the radiation is emitted from the $y=0$ level and escapes by the Doppler shift effect.

(iii) *The Polarization of the Radiation*

Radiation emitted by the Schwinger process should be plane polarized if the magnetic field concerned were everywhere uniform in direction. In general, however, it would be randomly polarized. The radiation is emitted from regions of low magnetic field (see (iv) below) and from levels in the solar atmosphere (§ 2) where the dielectric constant should not depart greatly from unity. It seems very unlikely, therefore, that the radiation reaching the earth should show appreciable signs of circular polarization. In the case of radiation from the $y=0$ level this should be strongly circularly polarized in a direction corresponding to the extraordinary ray.

Most enhanced radiation observed was strongly circularly polarized and if the generally observed law of sunspot polarity held for the spots in question, then the radiation received corresponded to the ordinary ray. Bursts and outbursts are usually randomly polarized but the latter are sometimes circularly polarized during part of their lifetime. It is concluded that the observed polarization of enhanced radiation is not consistent with an origin by the Schwinger mechanism and that none of the high intensity components have polarization consistent with an origin at the $y=0$ level.

(iv) *Regions of Origin of the Radiation*

The frequency at which maximum emission occurs by the Schwinger process is related to the magnetic field H by the equation $H=(2\boldsymbol{mc}/\boldsymbol{e})/\omega_m/\alpha^2$. If ω_m corresponds to 100 Mc/s, then $H=71/\alpha^2$ and substituting values of $\alpha=100$ for enhanced radiation and 10^5 for outbursts, the corresponding values of H are about 7×10^{-3} gauss and 7×10^{-9} gauss respectively. This suggests that if the emission were by this process then it would tend to take place in regions *well away* from large spots and not, as observed, within a solar radius or less of these spots (see, for example, Smerd 1950).

§ 6. DISCUSSION

The theory of high-intensity solar radiation in terms of thermal emission by electrons gyrating in a magnetic field has been tested by comparing observed and predicted characteristics of the radiation. Whether the radiation is presumed to escape by the Schwinger or Doppler shift mechanisms there is marked disagreement between theory and experiment. The discrepancies are so great that it appears necessary to discard both forms of the theory.

Before concluding it is desirable to point out two further objections, of a more general nature, which apply to any of the thermal theories discussed above. As evidence in favour of his thermal theory Ryle (1949 a) attempted to show that solar radiation by non-thermal processes such as those found in discharge tubes was not likely to be appreciable. He showed that certain characteristics of a discharge tube, notably a sharp plasma boundary, were unlikely to be duplicated in the solar atmosphere and concluded that radiation due to plasma oscillations would not exist. However, other workers (notably Bohm and Gross 1949) have pursued the investigation further and advanced sound reasons why solar radiation by non-thermal processes *should* be emitted. In this connection it is important to note that if *one* of the high-intensity components is found incapable of explanation in terms of purely thermal processes, then an effective non-thermal process must exist; this at once overcomes the difficulty of explaining the *other* high-intensity components.

A second general objection to the thermal theories lies in the proposed mechanism of obtaining the very fast electrons necessary. Both Ryle and Giovanelli suggest an origin in the electric fields induced by the movement of magnetic fields in the atmospheric gases. This mechanism has recently been criticized by the author (at the 1952 Sydney conference of Union Radio Scientifique Internationale) on the grounds that the gases in question move with the magnetic fields and have no appreciable electric fields induced in them.

It is concluded that none of the high-intensity components of solar radiation may be satisfactorily explained in terms of purely thermal process; ordered (non-thermal) motion of electrons must play a part.

Finally, it may be pointed out that Ryle's (1949 b) theory of cosmic radio waves (and of cosmic rays) by thermal processes, is open to similar criticism. Brightness temperatures postulated by him of 10^{14} deg K or more, if they really occur, are as difficult to explain as those of 10^{10} deg K on the Sun. In any case if non-thermal radiation originates on the Sun it is likely to do so elsewhere in the universe and to provide a very reasonable source of cosmic radio-frequency radiation.

ACKNOWLEDGMENTS

The author is grateful to Dr. J. L. Pawsey and Dr. D. F. Martyn for assistance with the manuscript.

REFERENCES

ALFVÉN, H., and HERLOFSON, N., 1950, *Phys. Rev.*, **78**, 616.
BOHM, D., and GROSS, E. P., 1949, *Phys. Rev.*, **75**, 1864.
DENISSE, J. F., 1947, *Ann. Astrophys.*, **10**, 1.
GIOVANELLI, R. G., 1948, *Nature, Lond.*, **161**, 133.
KIEPENHEUER, K. O., 1946, *Nature, Lond.*, **158**, 340.
MARTYN, D. F., 1946, *Nature, Lond.*, **158**, 632.
PAWSEY, J. L., 1950, *Proc. Instn. Elect. Engrs.*, **97**, 290.
PAWSEY, J. L., and YABSLEY, D. E., 1949, *Aust. J. Sci. Res.* A, **2**, 198.
PAYNE-SCOTT, R., and LITTLE, A. G., 1951, *Aust. J. Sci. Res.* A, **4**, 508; 1952, *Ibid.* A, **5**, 32.
PIDDINGTON, J. H., and MINNETT, H. C., 1951, *Aust. J. Sci. Res.* A, **4**, 131.
RYLE, M., 1948, *Proc. Roy. Soc.* A, **195**, 82; 1949 a, *Proc. Phys. Soc.* A, **62**, 483; 1949 b, *Ibid.*, A, **62**, 491; 1950, *Rep. Prog. Phys.*, **13**, 184 (London : Physical Society).
SAHA, M. N., 1946, *Nature, Lond.*, **158**, 549.
SAHA, M. N., BANNERJEA, B. K., and GUHA, U. C., 1947, *Indian J. Phys.*, **21**, 199.
SCHWINGER, J., 1949, *Phys. Rev.*, **75**, 1912.
SMERD, S. F., 1950, *Proc. Instn. Elect. Engrs.*, **97**, 447.
WALDMEIER, M., and MULLER, H., 1950, *Z. Astrophys.*, **27**, 58.
WESTFOLD, K. C., 1951, *J. Atmos. Terr. Phys.*, **1**, 152.
WILD, J. P., 1950 a, *Aust. J. Sci. Res.* A, **3**, 399; 1950 b, *Ibid.* A, **3**, 541; 1951, *Ibid.* A, **4**, 36.

Some Studies on Random Fading Characteristics

By R. B. BANERJI

University College of Science and Technology, Institute of Radiophysics and Electronics, University of Calcutta

Communicated by S. K. Mitra; MS. received 15*th April* 1952, *and in final form* 21*st October* 1952

ABSTRACT. The theory of random signals, as applied by various workers to the phenomenon of random fading of pulses singly reflected from the ionosphere, is considered critically. It is shown that, contrary to the prevalent idea, the velocity distribution of fading is independent of the power spectrum of the returned wave. This invalidates the ordinarily accepted explanation of certain observed departures from theory. An alternative explanation of these departures is given, which explanation is quantitative in character and can therefore be experimentally verified. The influence of a steady signal superposed on the randomly fading wave, on the velocity distribution of fading is investigated. It is found that unlike the effect on the amplitude distribution, the effect in this case is very small, resulting only in a very slight increase in the velocity of fading. An alternative method of determining the autocorrelation function of the fading pattern (from which the power spectrum of the returned wave can be derived by inverse Fourier transform) is given. This provides an alternative method of determining the velocity of drift of ionospheric irregularities. The method of deriving the theoretical power spectrum returned by drifting ionospheric irregularities has been improved upon, yielding a form of the power spectrum which appears to be more in accordance with observations.

§ 1. INTRODUCTION

RADIO waves received by reflection from the ionosphere are found to undergo spontaneous variations of amplitude, commonly known as 'fading'. It has been found that the fading variations can be periodic, quasi-periodic or random. Often, one type of fading is superimposed upon another. These have been ascribed to various causes. For instance, periodic or quasi-periodic fadings may be caused by interference between the ground wave and the sky wave or between sky waves reflected from different regions of the ionosphere (Banerji and Mukherji 1948, Banerji and Singh 1949). Similar types of fading may also be observed when there is interference between the two magneto-ionically split components. However, if observations be made on pulses singly reflected from the ionosphere (as in ionospheric exploration) instead of on a continuous wave or on broadcast transmission, then the first of the above two causes becomes inoperative and any periodic fading observed is only to be ascribed to interference between insufficiently resolved magneto-ionic components (Mitra 1949 a, 1950). Now, it has been found that fading occurs even when only one of the magneto-ionic components is received (by circularly polarized aerials), that is, even with a single downcoming wave. Such fadings are of a random character and have been ascribed by Ratcliffe (1948) to the presence of irregularities in the reflecting or transmitting regions of the ionosphere. These irregularities act as scattering centres and there is interference between the large number of incoherently scattered echoes. Some aspects of this type of random fading will be considered in this paper. In particular the following points will be discussed.

(i) *A neglected term in Fürth and MacDonald's formula for random noise* (§ 2). Since the scattering centres are always subject to motions, slight Doppler shifts will be introduced in the frequency of the scattered waves. The composite echo received may then be looked upon as due to interference between a number of waves whose frequencies are spread over a narrow power spectrum and which are in random phase. The phenomenon is thus analogous to random circuit noise which has been analysed by Fürth and MacDonald (1947). S. N. Mitra (1949 b) and McNicol (1950) have, therefore, utilized the results of Fürth and MacDonald for the determination of the random motion of the scattering centres from observations on the fading of singly reflected pulses. Both the authors have, however, reported a significant departure of the velocity of fading from the gaussian distribution predicted by Fürth and MacDonald. Mitra has tried to explain this departure on the assumption that the velocity distribution among the random scatterers is different from the gaussian, and hence gives rise to a non-gaussian power spectrum. It can, however, be shown that the shape of the velocity distribution curve will be gaussian irrespective of the shape of the power spectrum of the returned wave. There must, therefore, be some explanation other than that given by Mitra of the departures observed by him. It will be shown that the origin of the departure is traceable to an approximation used by Fürth and MacDonald which, though of little importance in their experimental condition, gives rise to a significant departure under Mitra's or McNicol's condition of experiment.

(ii) *Effect of the presence of a steady component on the shape of the velocity distribution of fading* (§§ 3 and 4). The effect of a steady component on the velocity distribution of fading does not appear to have been studied, though its effect on the amplitude distribution has been analysed by Rice (1945). The latter effect consists of a definite deviation in the shape from the amplitude distribution deduced by Rayleigh. A need for the study of the former is, however, also there, since a steady component in the returned echo has been found by McNicol (1950) and also by Banerji—rather strong—when the echo is scattered from the irregular structure of the E_s region of the ionosphere (1951). As rapid random fading generally coincides with the incidence of the E_s, it is pertinent to enquire what the shape of the velocity distribution would be in the presence of a steady component. The analysis reveals that the deviation from gaussian is rather small, the change being towards a slight increase in the velocity of fading.

(iii) *A possible method of investigating the motion of the ionospheric irregularities* (§§ 5 and 6). As pointed out above, the shape of the velocity distribution curve does not give us any idea of the shape of the power spectrum of the components of the returned composite echo. A method has been described whereby, by studying the variation of the standard deviation of the distribution with the interval chosen for the calculation of the velocity of fading, an insight is obtained into the nature of the power spectrum. However, to make any use of this knowledge of the power spectrum we must have a knowledge of the power spectra due to different types of motion in the ionospheric irregularities. We consider two extreme types of motion, namely, a steady drift and an absolutely random motion of the irregularities. Ratcliffe (1948) has considered the power spectrum of the case in which the echoes are from irregularities in absolutely random motion. It is recognized, however, that steady drifts of the irregularities are also very often present. We have, therefore, deduced in § 6 the shape of the power spectrum for the case of an ionosphere containing steadily drifting irregularities. By comparing the power spectrum obtained by the above analysis with those deduced from drifting and random models we can distinguish quantitatively between the two types of motion.

Results of closer study into the effect of different power spectra on fading curves and methods for their analysis will be published in the near future. The effect of the superposition of the two types of motion is at present under investigation.

§ 2. RECONSIDERATION OF FÜRTH AND MACDONALD'S THEORY AND SOME OF ITS CONSEQUENCES

Fürth and MacDonald start with the expression for the joint distribution formula given by Rice (1945). According to this formula, the joint probability $dp_\tau(R_1, R_2)$ that the amplitudes of a randomly varying signal are R_1 and R_2 at the instants t and $t+\tau$ respectively is given by

$$dp_\tau(R_1, R_2) = \frac{R_1R_2}{A} I_0\left\{\frac{R_1R_2}{A}(\mu^2+\lambda^2)^{1/2}\right\}\exp\left\{-\frac{\psi}{2A}(R_1{}^2+R_2{}^2)\right\} dR_1\, dR_2 \quad ..(2.1)$$

where
$$A = \psi^2-(\mu^2+\lambda^2),\ \psi = \int W(\nu)\,d\nu,\ \mu = \int W(\nu)\cos 2\pi(\nu-\nu_0)\tau\, d\nu,$$
$$\lambda = \int W(\nu)\sin 2\pi(\nu-\nu_0)\tau\, d\nu. \quad ......(2.2)$$

$W(\nu)$ represents the power spectrum and ν_0 the midband frequency of the returned wave. This general expression has been used by Fürth and MacDonald for the case of a train of waves having a narrow gaussian power spectrum. We shall apply the reasonings of these authors to the case of any power spectrum. If we put $(\mu^2+\lambda^2)^{1/2} = \psi(1-\epsilon)$, $R_1/\sqrt{\psi} = r_1$, $R_2/\sqrt{\psi} = r_2$ eqn. (2.1) may be written

$$dp_\tau = \frac{r_1r_2}{\epsilon(2-\epsilon)} I_0\left(\frac{1-\epsilon}{\epsilon}\frac{r_1r_2}{2-\epsilon}\right)\exp\left\{-\frac{r_1{}^2+r_2{}^2}{2\epsilon(2-\epsilon)}\right\} dr_1\, dr_2. \quad ......(2.3)$$

This, again, we can write as

$$dp_\tau = \frac{r_1r_2}{\epsilon(2-\epsilon)}\exp\left\{-\frac{r_1r_2}{\epsilon(2-\epsilon)}\right\} I_0\left(\frac{1-\epsilon}{\epsilon}\frac{r_1r_2}{2-\epsilon}\right)\exp\left\{-\frac{(r_1-r_2)^2}{2\epsilon(2-\epsilon)}\right\} dr_1\, dr_2. \quad ....(2.4)$$

Introducing the 'velocity' of fading v_τ defined as $v_\tau = r_1-r_2$ putting $r_2(r_2+v_\tau) = z^2$ and integrating over z from 0 to ∞ we get the velocity distribution

$$dp(v_\tau) = \frac{\exp\{-v_\tau{}^2/2\epsilon(2-\epsilon)\}dv_\tau}{\{2\pi\epsilon(2-\epsilon)(1-\epsilon)\}^{1/2}}\int_0^\infty \frac{2\{2\pi(1-\epsilon)\}^{1/2}}{\{\epsilon(2-\epsilon)\}^{1/2}} I_0\left\{\frac{1-\epsilon}{\epsilon}\frac{z^2}{2-\epsilon}\right\}$$
$$\times\frac{z^3\exp\{-z^2/\epsilon(2-\epsilon)\}}{\{v^2+4z^2\}^{1/2}}dz. \quad ......(2.5)$$

To integrate this analytically Fürth and MacDonald make the assumption that ϵ is sufficiently small (i.e. the power spectrum is narrow and the time interval τ small) to warrant the replacement of $I_0(x)$ by its asymptotic expansion $e^x/(2\pi x)^{1/2}$ so that (2.5) reduces to

$$dp(v_\tau) = \frac{\exp\{-v_\tau{}^2/2\epsilon(2-\epsilon)\}dv_\tau}{\{2\pi\epsilon(2-\epsilon)(1-\epsilon)\}^{1/2}}\int_0^\infty \frac{2z^2\exp\{-z^2/(2-\epsilon)\}}{(v_\tau{}^2+4z^2)^{1/2}}dz. \quad ......(2.6)$$

The assumption that ϵ is small makes another simplification possible. The integral is a rather slowly decreasing function of v_τ. Hence, if ϵ is small enough to make the exponential term in (2.6) very sharp, any effect of the variation of the integral will be suppressed and the effect will be as if the integral has its value at the origin for all values of v_τ, i.e. that

$$dp(v_\tau)=\frac{\exp\{-v_\tau^2/4\epsilon\}}{(4\pi\epsilon)^{1/2}}\,dv_\tau. \qquad \ldots\ldots(2.7)$$

It will be noticed that except for the narrowness no assumption has been made regarding the form of the spectrum, and the gaussian form of the velocity distribution is not a consequence of the power spectrum being gaussian.

The explanation given by Mitra needs a detailed discussion in this connection. According to that author, the deviation of his experimental curve from the gaussian is due to the fact that the distribution of velocities among the scattering sources, and hence the power spectrum of the returning wave, is different from gaussian. It may, however, be noted that the deduction of Fürth and MacDonald's theoretical result as analysed above, shows that a gaussian distribution of the velocity of fading will result irrespective of the shape of the above power spectrum. It is only the standard deviation of the velocity distribution and not its actual shape which is determined by the power spectrum. The deviations observed by Mitra are therefore not adequately explained by his theory.

One may enquire if the deviations are due to the assumption that ϵ is small. In Fürth and MacDonald's experiment on radio noise, τ was taken to be of the order of 10^{-3} second, a fact that justified their assumption, and hence their experimental results agreed with theory. In the case of fading analysis, however, τ is about 100 times larger, and it is doubtful whether corresponding values of ϵ would be small enough to justify Fürth and MacDonald's approximations.

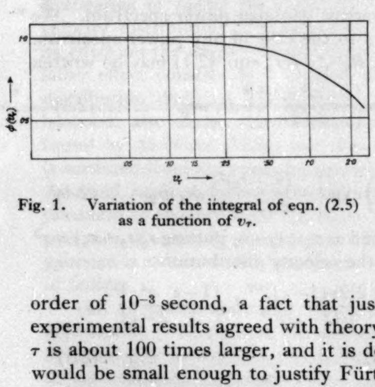


Fig. 1. Variation of the integral of eqn. (2.5) as a function of v_τ.

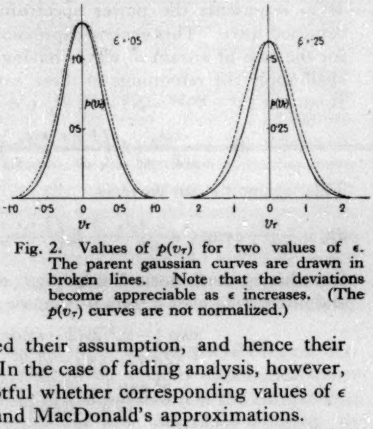


Fig. 2. Values of $p(v_\tau)$ for two values of ϵ. The parent gaussian curves are drawn in broken lines. Note that the deviations become appreciable as ϵ increases. (The $p(v_\tau)$ curves are not normalized.)

To investigate this point we have integrated numerically the integral in eqn. (2.5) for different values of v_τ. Figure 1 shows the function obtained in this manner for $\epsilon=0{\cdot}10$. The form of the function is not very sensitive to changes in ϵ within the range of values likely to be encountered in actual fading study. Figure 2 shows the form of $p(v_\tau)$ for two different values of ϵ together with the nearest gaussian distributions. The increase of deviation with increasing ϵ may be noted.

§ 3. FADING CHARACTERISTICS IN PRESENCE OF A STEADY ECHO

To deduce the fading characteristics of a randomly varying wave in the presence of a steady component we have first to deduce the bivariate probability distribution of the amplitudes at instants separated by finite intervals of time.

The expression for this probability function, in the absence of a steady component, has been deduced by Rice and is given in eqn. (2.1) above. We will follow a procedure analogous to that followed by him.

According to Rice, a randomly varying signal can be written as

$$I(t) = I_c \cos 2\pi\nu_0 t + I_s \sin 2\pi\nu_0 t \qquad \ldots\ldots(3.1)$$

where I_c and I_s are normally distributed and ν_0 is the midband frequency of the spectrum of the fluctuating wave form. $R = (I_c^2 + I_s^2)^{1/2}$ is easily seen to give the instantaneous amplitude. If we are to deal with the signal at two instants t and $t+\tau$ then we have to consider the probability distribution of four variables I_{c1}, I_{s1}, I_{c2}, I_{s2} related to $I(t)$ and $I(t+\tau)$ by

$$I(t) = I_{c1} \cos 2\pi\nu_0 t + I_{s1} \sin 2\pi\nu_0 t, \qquad \ldots\ldots(3.2\,a)$$

$$I(t+\tau) = I_{c2} \cos 2\pi\nu_0(t+\tau) + I_{s2} \sin 2\pi\nu_0(t+\tau). \qquad \ldots\ldots(3.2\,b)$$

In this case, the tetravariate probability distribution of the I's is given by

$$dp_\tau = \frac{dI_{c1}\, dI_{s1}\, dI_{c2}\, dI_{s2}}{4\pi^2(\psi^2 - \mu^2 - \lambda^2)} \exp\left[-\frac{1}{2(\psi^2-\mu^2-\lambda^2)}\{\psi(I_{c1}^2 + I_{s1}^2 + I_{c2}^2 + I_{s2}^2) - 2\mu(I_{c1}I_{c2} + I_{s1}I_{s2}) - 2\lambda(I_{c1}I_{s2} - I_{c2}I_{s1})\}\right] \qquad \ldots\ldots(3.3)$$

ψ, μ, λ being defined in eqn. (2.2).

If again, the signal be not purely random, but has a separate steady signal of amplitude P, then the signals at t and $t+\tau$ are to be written as

$$I(t) = (P + I_{c1}) \cos 2\pi\nu_0 t + I_{s1} \sin 2\pi\nu_0 t, \qquad \ldots\ldots(3.4\,a)$$

$$I(t+\tau) = (P + I_{c2}) \cos 2\pi\nu_0(t+\tau) + I_{s2} \sin 2\pi\nu_0(t+\tau). \qquad \ldots\ldots(3.4\,b)$$

In this case, the four components are not completely random in the sense that their mean values are no more zero and their product moments are much increased. By defining the corresponding amplitudes by R_1 and R_2 the completely random variables now can be expressed as

$$R_1 \cos\theta_1 - P = I_{c1},\; R_1 \sin\theta_1 = I_{s1},\; R_2 \cos\theta_2 - P = I_{c2},\; R_2 \sin\theta_2 = I_{s2} \qquad \ldots(3.5)$$

the θ's being arbitrarily introduced parameters to retain $R_1^2 = (P + I_{c1})^2 + I_{s1}^2$, and similarly for R_2.

We have now to express eqn. (3.3) as a function of the R's so that integration over θ_1 and θ_2 will give us the joint probability distribution of R_1 and R_2. It will simplify calculations if we assume that the frequency spectrum of the signal is symmetrical. This is so in most of the cases of waves returned from an irregularly reflecting ionosphere. Hence we can put $\lambda = 0$ so that replacing the values of I_{c1}, I_{s2}, etc. in (3.3) we have

$$dp_\tau = \frac{R_1R_2}{4\pi^2(\psi^2-\mu^2)} \exp\left[-\frac{P^2}{\psi+\mu} - \frac{\psi(R_1^2+R_2^2)}{2(\psi^2-\mu^2)} + \frac{PR_1\cos\theta_1}{\psi+\mu} + \left\{\left(\frac{PR_2}{\psi+\mu} + \frac{\mu R_1R_2}{\psi^2-\mu^2}\cos\theta_1\right)\cos\theta_2 + \frac{\mu R_1R_2}{\psi^2-\mu^2}\sin\theta_1\sin\theta_2\right\}\right] dR_1\, dR_2\, d\theta_1\, d\theta_2. \qquad \ldots\ldots(3.6)$$

We first integrate over θ_2. The expression for dp_τ can then be written as

$$dp_\tau = \frac{R_1R_2}{2\pi(\psi^2-\mu^2)} \exp\left[-\frac{P^2}{\psi+\mu} - \frac{\psi(R_1^2+R_2^2)}{2(\psi^2-\mu^2)}\right] \exp\left[\frac{PR_1}{\psi+\mu}\cos\theta_1\right] \times I_0\left[\left\{\left(\frac{PR_2}{\psi+\mu}\right)^2 + \left(\frac{\mu R_1R_2}{\psi^2-\mu^2}\right)^2 + 2\frac{PR_2\mu R_1R_2}{(\psi+\mu)(\psi^2-\mu^2)}\cos\theta_1\right\}^{1/2}\right] dR_1\, dR_2\, d\theta_1 \qquad \ldots\ldots(3.7)$$

since $\frac{1}{2\pi}\int_{\varphi}^{2\pi+\varphi} \exp(iz\cos\theta)\,d\theta = J_0(z)$ and $J_0(iz)=I_0(z)$ (Watson 1944, pp. 20, 77).

This has now to be further integrated with respect to θ_1. For this we recall that

$$I_0\{(a^2+b^2+2ab\cos\theta)^{1/2}\}=I_0(a)I_0(b)+2\sum_{n=1}^{\infty} I_n(a)I_n(b)\cos n\theta^*$$

whence (3.7) can be written as

$$dp_\tau(R_1, R_2)=\frac{R_1R_2\,dR_1\,dR_2}{\psi^2-\mu^2}\exp\left\{-\frac{P^2}{\psi+\mu}-\frac{\psi(R_1{}^2+R_2{}^2)}{2(\psi^2-\mu^2)}\right\}I_0\left[\frac{\mu R_1R_2}{\psi^2-\mu^2}\right]$$
$$\times\left\{I_0\left(\frac{PR_1}{\psi+\mu}\right)I_0\left(\frac{PR_2}{\psi+\mu}\right)+2\sum_{n=1}^{\infty}I_n\left(\frac{PR_1}{\psi+\mu}\right)I_n\left(\frac{PR_2}{\psi+\mu}\right)\frac{I_n\{\mu R_1R_2/(\psi^2-\mu^2)\}}{I_0\{\mu R_1R_2/(\psi^2-\mu^2)\}}\right\}. \qquad \ldots\ldots(3.8)$$

It will be noticed that eqn. (3.8) reduces to eqn. (2.1) if $P=0$, as expected.

§ 4. THE SPEED OF FADING

We will now deduce the velocity distribution of fading by a method analogous to that of Fürth and MacDonald. Before that, however, a few approximations have to be introduced. These approximations follow basically from that of Fürth and MacDonald, namely that the spectrum is narrow or $(\psi-\mu)/\psi$ is small.

Under such conditions we can take $\mu R_1R_2/(\psi^2-\mu^2)$ to be large except for very small values of R_1 and R_2. Hence, since

$$I_n(z)\simeq I_0(z)\simeq e^z/(2\pi z)^{1/2} \qquad \ldots\ldots(4.1)$$

eqn. (3.8) can be simplified. We will first make substitutions similar to those in § 2, and write $(\psi-\mu)/\psi=\epsilon$; $R_1/\sqrt{\psi}=r_1$; $R_2/\sqrt{\psi}=r_2$; $P/\sqrt{\psi}=q$ so that after the above approximation, we have

$$dp_\tau=\frac{r_1r_2}{(4\pi\epsilon)^{1/2}(1-\epsilon)/(1-\frac{1}{2}\epsilon)}\exp\left(-\frac{q^2}{2-\epsilon}\right)\exp\left\{-\frac{(r_1-r_2)^2}{2\epsilon(2-\epsilon)}\right\}\exp\left(-\frac{r_1r_2}{2-\epsilon}\right)$$
$$\times(r_1r_2)^{1/2}I_0\left\{\frac{q}{2-\epsilon}(r_1+r_2)\right\}dr_1\,dr_2. \qquad \ldots\ldots(4.2)$$

Defining the velocity of fading v_τ as in § 2, transforming the coordinates to r_2 and v_τ putting $v_\tau{}^2+4r_2(r_2+v_\tau)=x^2$ and integrating over x

$$dp(v_\tau)=\frac{\exp\{-q^2/(2-\epsilon)\}dv_\tau}{(4\pi\epsilon)^{1/2}(1-\epsilon)/(1-\frac{1}{2}\epsilon)}\exp\left\{-\frac{v_\tau{}^2}{2\epsilon(2-\epsilon)}+\frac{v_\tau{}^2}{4(2-\epsilon)}\right\}$$
$$\times\frac{1}{4}\int_{v_\tau}^{\infty}I_0\left(\frac{qx}{2-\epsilon}\right)\exp\left\{-\frac{x^2}{4(2-\epsilon)}\right\}(x^2-v_\tau{}^2)^{1/2}\,dx. \qquad \ldots\ldots(4.3)$$

Since we need this integral only for small value of v_τ (large values being taken care of by the strong exponential term in (4.3)) we put $(x^2-v_\tau{}^2)^{1/2}\simeq x$ so that the integral becomes

$$\int_{v_\tau}^{\infty}I_0\left(\frac{qx}{2-\epsilon}\right)x\exp\left\{-\frac{x^2}{4(2-\epsilon)}\right\}dx \qquad \ldots\ldots(4.4)$$

and, since for small values of x, $I_0(x)\simeq 1$, the integral is approximately

$$\left[\int_0^{\infty}I_0\left(\frac{qx}{2-\epsilon}\right)x\exp\left\{-\frac{x^2}{4(2-\epsilon)}\right\}dx-\int_0^{v_\tau}x\exp\left\{-\frac{x^2}{4(2-\epsilon)}\right\}dx\right]. \qquad \ldots\ldots(4.5)$$

---

* Watson (1944) p. 128, remembering $J_n(iz)=i^n I_n(z)$ and $J_n(-z)=(-1)^n J_n(z)$.

The first integral can be shown (Watson 1944, p. 393) to be equal to

$$2(2-\epsilon)\exp\{q^2/(2-\epsilon)\} \qquad \ldots\ldots(4.6\,a)$$

while the second easily simplifies to

$$2(2-\epsilon)[1-\exp\{-v_\tau{}^2/4(2-\epsilon)\}]. \qquad \ldots\ldots(4.6\,b)$$

Equation (4.3) therefore takes the form

$$dp(v_\tau) = \frac{(1-\frac{1}{2}\epsilon)^2}{(4\pi\epsilon)^{1/2}(1-\epsilon)}\exp\left[-\frac{v_\tau{}^2}{2\epsilon(2-\epsilon)}+\frac{v_\tau{}^2}{4(2-\epsilon)}\right]$$
$$\times\left[1-\exp\left\{-\frac{q^2}{2-\epsilon}\right\}+\exp\left\{-\frac{q^2}{2-\epsilon}-\frac{v_\tau{}^2}{4(2-\epsilon)}\right\}\right] \qquad \ldots\ldots(4.7)$$

giving us the complete expression. It is to be noticed that the result does not take into account the correction term introduced in § 2, and hence is accurately valid for only small values of ϵ. It will also be noticed that the curve remains gaussian both for very small and very large values of q, being slightly wider in the latter case. The difference between the two extreme gaussian distributions being small, it is evident that for intermediate values of q the difference will also be small. Figure 3 indicates two such curves and the nearest gaussian distributions.

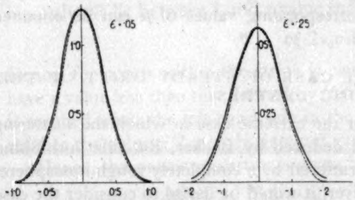


Fig. 3. Velocity distribution of fading in the presence of a steady signal of strength twice the r.m.s. value of the random signal ($q=2$). The corresponding curves without the steady signal are shown in broken lines alongside each curve. It will be noticed that the widening effect of the steady signal is greater the greater the value of ϵ. (The curves are not normalized.)

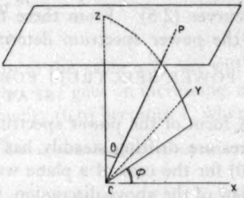


Fig. 4. Disposition of a typical scattered ray in relation to the position of the observer and the scattering ionospheric plane.

§ 5. DERIVATION OF THE SHAPE OF THE POWER SPECTRUM FROM FADING DATA

It has been pointed out in § 2 that the shape of the power spectrum of the returned wave cannot be deduced from the delineation of the velocity distribution of fading. Some knowledge about the nature of the shape of the frequency spectrum is, however, very useful, because from such knowledge the nature of the motion of the ionospheric irregularities can be known. Hence, we discuss a method whereby the shape of the power spectrum can be deduced by noting how the standard deviation of the velocity distribution changes as one changes the duration τ on which our analysis has been based.

We have, from definition

$$\mu(\tau)=\int_0^\infty W(\nu)\cos 2\pi(\nu-\nu_0)\tau\,d\nu;\quad \lambda(\tau)=\int_0^\infty W(\nu)\sin 2\pi(\nu-\nu_0)\tau\,d\nu, \qquad \ldots(5.1)$$

whence

$$\mu+i\lambda=\int_0^\infty W(\nu)\exp\{2\pi i(\nu-\nu_0)\tau\}\,d\nu. \qquad \ldots\ldots(5.2)$$

From this it follows from the theory of Fourier transform that

$$W(\nu)=\int_0^\infty (\mu+i\lambda)\exp\{-2\pi i(\nu-\nu_0)\tau\}\,d\tau \qquad \ldots\ldots(5.3)$$

the real part of which gives us

$$W(\nu)=\int_0^\infty \mu(\tau)\cos 2\pi(\nu-\nu_0)\tau\,d\tau+\int_0^\infty \lambda(\tau)\sin 2\pi(\nu-\nu_0)\tau\,d\tau. \qquad \ldots\ldots(5.4)$$

Since in almost all practical cases the returned power spectrum is symmetrical we can put $\lambda=0$. Hence, if we know μ as a function of τ, we can find $W(\nu)$ by the Fourier transform (5.4). We therefore give τ a number of successive values and analyse the corresponding fading curves to get a number of values of ϵ in the curves (2.5). From these the corresponding values of μ can be obtained and the power spectrum determined.

§ 6. POWER SPECTRUM FOR THE CASE OF STEADY DRIFT OF THE SCATTERING CENTRES

A form of the power spectrum for the extreme case in which the scattering centres are drifting steadily has been deduced by Booker, Ratcliffe and Shinn (1950) for the case of a plane wave scattered by a completely rough ionosphere. In view of the above discussion, however, it would be useful to consider the case of similar scattering of a spherical wave which is more akin to the radiation patterns of the actually employed transmitting aerials.

In fig. 4 let ZP be the plane of the ionosphere, with the scattering centres drifting in the X direction with velocity $-v$. Let h be the height of ZP above the observation point O. Then any point P on the plane can be represented by the spherical polar coordinates (θ, ϕ). We assume that the aerial has an omnidirectional characteristic, that the scattering centres are uniformly distributed over the plane and that the coefficient of back scatter is the same for all the centres. Then the power scattered back to unit area at a given point on the ground by an element $d\omega$ of the ionosphere is directly proportional to the solid angle subtended by the element at O and by that subtended by unit area on ground at the element, i.e. to $d\omega \cos^2\theta/r^2$.

Let A be the power scattered back per unit area overhead, i.e. where the solid angle is $1/h^2$. Then the power scattered back from any element of area $d\omega$ is seen to be, on replacing values of $d\omega$ and r

$$W(\theta,\phi)=Ah^2\sin\theta\cos\theta\,d\theta\,d\phi. \qquad \ldots\ldots(6.1)$$

Now, the Doppler shift ξ undergone by the ray scattered back from the point (θ, ϕ) is dependent on (θ, ϕ) and is easily seen to be given by

$$\xi=-(2\nu_0 v/\mathbf{c})\sin\theta\cos\phi. \qquad \ldots\ldots(6.2)$$

To find the total power scattered from the points within an area in the plane in which ξ is constant, we transform coordinates according to the following scheme

$$\cos\phi = (\mathbf{c}\,\mathrm{cosec}\,\Theta/2\nu_0 v)\xi = y, \text{ say } \theta = \Theta \qquad \ldots\ldots(6.3)$$

and integrate over the elements of constant ξ; $d\theta\, d\phi$ will transform according to the scheme

$$d\theta\, d\phi = \begin{vmatrix} 0 & 1 \\ \dfrac{c\,\mathrm{cosec}\,\Theta/2\nu_0 v}{(1-y^2)^{1/2}} & \dfrac{c\cot\Theta\,\mathrm{cosec}\,\Theta/2\nu_0 v}{(1-y^2)^{1/2}} \end{vmatrix}\, d\Theta\, d\xi$$

where y is given by (6.3), so that we have, transforming (6.1)

$$W(\Theta, \nu) = Ah^2 \frac{\mathbf{c}\cos\Theta}{2\nu_0 v} \cdot \frac{d\nu\, d\Theta}{(1-y^2)^{1/2}}. \qquad \ldots\ldots(6.4)$$

Integrating over Θ

$$W(\nu) = \frac{Ah^2\mathbf{c}}{\nu_0 v} \int \frac{\cos\Theta\, d\Theta}{(1-y^2)^{1/2}}. \qquad \ldots\ldots(6.5)$$

The limits of this integration are determined from the following consideration. For a given ξ it is easily seen that the points producin gthis amount of Doppler shift are confined within a strip of the plane in which θ and ϕ are connected by eqn. (6.3). It will be seen that for values of ϕ lying between 0 and 2π, cosec Θ must always lie between 1 and a value such that

$$\mathbf{c}\xi/2\nu_0 v < \sin\Theta. \qquad \ldots\ldots(6.6)$$

Hence, for points within the strips Θ can go up to the value $\pi/2$, but will never have a value less than that defined by (6.6). Also, as ξ goes on increasing, a value is reached when the left-hand side in the inequality (6.6) becomes 1, whereby Θ reaches the value $\pi/2$ and the integral vanishes.

Hence, the power spectrum is given by

$$W(\nu) = \frac{Ah^2\mathbf{c}}{\nu_0 v}\left[1 - \left\{\frac{\mathbf{c}(\nu-\nu_0)}{2\nu_0 v}\right\}^2\right]^{1/2} \quad \text{for } |\nu-\nu_0| < 2\nu_0 v/\mathbf{c} \quad \ldots\ldots(6.7\,a)$$

and

$$W(\nu) = 0 \quad \text{for } |\nu-\nu_0| \geqslant 2\nu_0 v/\mathbf{c} \quad \ldots\ldots(6.7\,b)$$

ν_0 being the midband frequency which, evidently, is the same as the sounding frequency.

The power spectrum is thus seen to be semi-elliptical. Kerr and Shain (1951) have deduced a semi-circular power spectrum for a similar phenomenon, namely, fading radio waves scattered from the librating moon.

§7. CONCLUDING REMARKS

The discussions in the previous sections show that the prevalent idea that the velocity distribution of fading is the same as the power spectrum of the fading wave is untenable. It is, however, on this assumption that Mitra had offered explanation of the observed departures of the random fading curves from theory. The alternative explanation as provided here is independent of such assumption. Further, this explanation, unlike the previous one, is quantitative in character and makes experimental verification possible. Unfortunately, the published data and curves of random fading are not sufficient for the verification of the alternative explanation. An experimental set up for the necessary observations is shortly expected in this laboratory. When this materializes a quantitative verification of the suggested explanation will be possible.

The analysis also shows that, if the power spectrum be narrow, the shape of the velocity distribution is not very much affected by a steady reflection superposed on the randomly fading wave. There is only a slight increase in the velocity of fading. This result is in contrast to that obtained by Rice for the case of amplitude distribution where the superposition of a steady signal changes markedly the shape of the distribution curve.

An important parameter of a rough ionosphere is the velocity of the scattering centres. As has been indicated by Booker, Ratcliffe and Shinn (1950) the auto-correlation function of the fading curve can be utilized for determining this velocity. An alternative method has been suggested in this paper for determining the auto-correlation function from the variation of the standard deviation of the velocity distribution with the basic time interval τ. This is expected to yield the function more readily for small samples (Kendall 1946), being more closely related to the variate difference analysis. It has also been pointed out how the power spectrum of the returned wave is derivable as the inverse Fourier transform of the auto-correlation function.

The work of Booker, Ratcliffe and Shinn on the theoretical derivation of the power spectrum returned by steadily drifting ionospheric irregularities has been extended to the case of spherical waves. The interesting result obtained is that the power spectrum for the latter case gives rise to an oscillatory correlation function which is similar rather to that observed by McNicol than to that derived by Booker, Ratcliffe and Shinn.

The auto-correlation method provides a useful device for studying the motion of the ionospheric irregularities. Unlike Briggs, Phillips and Shinn's (1949) method however, the direction of the drift remains unknown in this method. The latter, however, may be obtained by suitably modifying the power spectrum by special design of the aerial system.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor S. K. Mitra for his constant encouragement and interest. Thanks are also due to the Ministry of Education, Government of India for having provided him with a scholarship under the Research Training Scheme as recommended by the Scientific Manpower Committee.

REFERENCES

Banerji, R. B., 1951, *Indian J. Phys.*, **25**, 359.
Banerji, S. S., and Mukherji, G. C., 1948, *Phil. Mag.*, **39**, 697.
Banerji, S. S., and Singh, R. N., 1949, *Sci. and Cult.*, **14**, 294.
Booker, H. G., Ratcliffe, J. A., and Shinn, D. H., 1950, *Phil. Trans. Roy. Soc.* A, **262**, 579.
Briggs, B. H., Phillips, G. J., and Shinn, D. H., 1949, *Proc. Phys. Soc.* B, **63**, 106.
Fürth, R., and MacDonald, D. K. C., 1947, *Proc. Phys. Soc.*, **59**, 388.
Kerr, F. J., and Shain, C. A., 1951, *Proc. Inst. Radio., Engrs.*, **39**, 230.
Kendall, M. G., 1946, *Contributions to the Study of Oscillatory Time Series* (Cambridge : University Press).
McNicol, R. W. E., 1950, *Proc. Instn. Elect. Engrs.*, **96**, Pt. III, 366.
Mitra, S. N., 1949 a, *Proc. Instn. Elect. Engrs.*, **96**, Pt. III, 441 ; 1949 b, *Ibid.*, **96**, Pt. III, 505 ; 1950, *Indian J. Phys.*, **24**, 197.
Ratcliffe, J. A., 1948, *Nature, Lond.*, **162**, 9.
Rice, S. O., 1945, *Bell System Tech. J.*, **25**, 48.
Watson, G. N., 1944, *A Treatise on the Theory of Bessel Functions*, 2nd edn. (Cambridge : University Press).

The Measurement of the Absolute Viscosity of Anomalous Fluids

I: The Measurement of the Time-dependence of Viscosity of Thixotropic Materials

By S. THORNTON

Department of Physics, Durham Colleges in the University of Durham

Communicated by W. A. Prowse ; MS. received 15th August 1952

ABSTRACT. The methods of measuring the absolute viscosity of anomalous fluids are reviewed briefly, and an improved method is given, which is an extension of Mooney's method, and which makes use of a simple concentric cylinder viscometer, It is shown that if a suitable technique is used information on the time-dependence of the absolute viscosity of thixotropic materials can be obtained either with this instrument or with a parallel disc viscometer.

---

§ 1. INTRODUCTION

THREE distinct methods have been suggested for determining the relation between the shearing stress and the velocity gradient in materials exhibiting anomalous flow properties (hereafter referred to as anomalous fluids), but the conditions of operation in each case are such that only anomalous fluids of one very special type can be used satisfactorily. Richardson and Tyler (1933) used anomalous fluids of the correct type and measured the shearing stress and the velocity gradient directly in a concentric cylinder viscometer but the method, which makes use of a velocity detecting probe, is open to the objection that the probe affects the flow condition in the sample. The second method, due to Lindquist and Sierichs (1950) who recognized the need for the correct type of anomalous fluid, makes use of a concentric cylinder viscometer which can be used with a number of cylinders of different radii so that an extrapolation can be made to zero gap width. Though fundamentally sound the method is inaccurate and does not compare with the third method which is potentially the best of the three and which alone will concern us here. It is based on the assumption that the shearing stress distribution is known at all points in the viscometer, and it is usually ascribed to Rabinowitsch (1929) or to Mooney (1931). The former developed it theoretically for the flow of anomalous fluids through a simple Poiseuille tube viscometer, and the latter extended it to the case of the concentric cylinder viscometer and considered also the effect of slip at the walls in both instruments. It has also been considered in slightly different forms by Schofield and Scott Blair (1930), Reiner (1930), Eisenschitz (1933), Søltoft (1948), Oldroyd (1949), Clark and Deutsch (1950) and Krieger and Maron (1952).

It is the purpose of this paper to show that the Mooney method can be developed, not only to enable measurements made with a concentric cylinder apparatus on a very special class of anomalous fluids to yield absolute viscosity determinations without recourse to approximate methods or to the use of extra cylinders (cf. Krieger and Maron), but also to enable a quantitative study of the time dependence of viscosity of thixotropic fluids to be undertaken, and this for the first time. Since a parallel disc viscometer is also well suited for these studies this too is discussed. The Poiseuille tube viscometer on the other hand is not considered here because it is suitable only for materials which show no time dependence of viscosity at constant shearing stress, and such materials, other than Newtonian fluids, exhibit characteristic rheological properties which tend to make the Rabinowitsch treatment invalid unless special precautions are taken. This particular aspect is considered in greater detail in a companion paper (Thornton and Rae 1953) which deals with the measurement of absolute viscosities in such an apparatus.

§ 2. CLASSIFICATION OF MATERIALS

For the purpose of this discussion it will be convenient to distinguish two types of anomalous fluids. Defining the viscosity as the ratio of the shearing stress to the rate of shear for a small element of volume of the fluid, the first is the normal thixotropic type, for which the viscosity changes are reversible and for which the viscosity is dependent on (*a*) the previous shear history of the element and (*b*) the applied shearing stress; the second is an instantaneously thixotropic type, for which the viscosity changes are reversible but for which the viscosity is dependent *only* on the applied shearing stress.

§ 3. CONDITIONING OF SAMPLES

Experimentally it is found that if a constant shearing stress f_0 is applied for a long time to a normal thixotropic sample the effect of previous shear history is obliterated and an equilibrium viscosity η_0, which is a function of the shearing stress only, is obtained. If subsequently a different shearing stress f is applied for a given time t the viscosity η at the end of that time is a function of f_0, f and t, and

$$\eta = F'(f_0, f, t). \qquad \ldots\ldots(1)$$

Similarly, since prolonged maintenance of a velocity gradient g_0 can also obliterate the effect of previous shear history, the subsequent maintenance of a velocity gradient g for a time t leads to the relationship

$$\eta = G(g_0, g, t). \qquad \ldots\ldots(2)$$

Thus if an element of volume of an anomalous fluid has been subjected first to the shearing stress f_0 till the equilibrium viscosity η_0 has been obtained, and then to the shearing stress f for the time t, an *effective* shear history which is reproducible has been obtained and, provided the viscosity η can be measured, a method is available for studying the time dependence of viscosity of thixotropic materials for particular values of f_0 and f. Alternatively if the velocity gradient is kept constant the time dependence of viscosity for particular values of g_0 and g can be investigated.

As will be shown in § 5 the absolute viscosity of thixotropic samples which have been conditioned in this way to the shearing stresses f_0 and f can be measured in a concentric cylinder viscometer, and in a parallel disc viscometer if the conditioning has been to shear rates g_0 and g.

§ 4. THE MEASUREMENT OF THE ABSOLUTE VISCOSITY OF INSTANTANEOUSLY THIXOTROPIC MATERIALS

(i) *The Concentric Cylinder Viscometer Method*

Consider an instantaneously thixotropic fluid in a concentric cylinder viscometer with inner and outer cylinder radii equal to a and b respectively. The torque C per unit length of the cylinder is given in the usual way by

$$C = 2\pi r^2 f \qquad \ldots\ldots(3)$$

where f is the shearing stress at a radial distance r and

$$f = -\eta r \, d\omega/dr \qquad \ldots\ldots(4)$$

if ω is the angular velocity of the fluid at the radial distance r. Thus since C is constant

$$d\omega = df/2\eta$$

and

$$\omega_{\alpha,\beta} = \int_{\beta}^{\alpha} df/2\eta \qquad \ldots\ldots(5)$$

if α and β are the shearing stresses at the radial distances a and b respectively (i.e. f_a and f_b), $\omega_{\alpha,\beta}$ is the relative angular velocity of the two cylinders, and there is no slip in the fluid at the walls. When b is infinite, that is for a semi-infinite sea viscometer,

$$\omega_{\alpha,0} = \int_{0}^{\alpha} df/2\eta \qquad \ldots\ldots(6)$$

gives the angular velocity of the inner cylinder if the outer is considered at rest, and

$$\eta_{\alpha} = \tfrac{1}{2}\, d\alpha/d\omega_{\alpha,0} \qquad \ldots\ldots(7)$$

where η_{α} is the viscosity at the shearing stress α. This gives

$$\eta_{\alpha} = \frac{1}{4\pi a^2}\frac{dC}{d\omega_{\alpha,0}} \qquad \ldots\ldots(8)$$

and the absolute viscosity can be determined from the slope of the graph relating C and $\omega_{\alpha,0}$. The procedure involves measuring the angular velocity and shearing couple of the inner cylinder of a semi-infinite sea viscometer, and is not practicable, but fortunately these values can be computed fairly simply from the data obtained with an ordinary concentric cylinder viscometer.

From eqn. (3) the ratio K of the shearing stress at the outer cylinder to that at the inner cylinder is always a constant for a given viscometer, and since $K = a^2/b^2 = \beta/\alpha$ it can be made as small as desirable by suitable choice of a and b.

Equation (5) for a practical viscometer can be rewritten

$$\omega_{\alpha,K\alpha} = \int_{K\alpha}^{\alpha} \frac{df}{2\eta}$$

and when there is no slip it can be shown that $\omega_{\alpha,0} = \omega_{\alpha,K\alpha} + \omega_{K\alpha,0}$ and $\omega_{K\alpha,0} = \omega_{K\alpha,K(K\alpha)} + \omega_{K(K\alpha),0}$ and so on, so that the angular velocity of the cylinder of a semi-infinite sea viscometer operating at a shearing stress α can be obtained by using a concentric cylinder viscometer with the inner cylinder operating at shearing stresses α, $K\alpha$, $K^2\alpha$ and so on, and summing the relative angular velocities between the inner and outer cylinders obtained in each case. Theoretically this is an infinite series, but in practice if K is equal to 0·1, the contribution of angular velocity from the third term is approximately only 1% of the whole, even with Newtonian fluids. With instantaneously thixotropic samples even the second term can be neglected in most cases.

Equation (8) can now be rewritten

$$\eta_\alpha = \frac{1}{4\pi a^2} \frac{dC}{d(\omega_{\alpha,\, K\alpha} + \omega_{K\alpha,\, K(K\alpha)} + \ldots)} \qquad \ldots\ldots(9)$$

and the few terms of the summation required to give a highly accurate result can readily be found from the graph of C plotted against $\omega_{\alpha,\, K\alpha}$.

(ii) *The Parallel Disc Viscometer Method*

Consider an instantaneously thixotropic fluid in a parallel disc viscometer with a guard ring to ensure uniform shearing between the two discs. Let the radius of the upper disc, normally supported from a calibrated torsion wire, be a and let the couple C' be applied to the upper when the lower disc rotates with an angular velocity ω.

For an annulus of the sample between radii r and $r+dr$ the velocity gradient g, if there is no slip, is everywhere $r\omega/h$, where h is the distance between the two discs, so that the couple C' is given by

$$C' = \int_0^a 2\pi r^2 (\eta g)\, dr$$

whence

$$\omega^3 C' = \tfrac{1}{2}\pi h^3 \int_0^\gamma \eta\, d(g^4)$$

where γ is the velocity gradient $a\omega/h$ at radius a and

$$\eta_\gamma = \frac{2}{\pi h^3} \frac{d(\omega^3 C')}{d\gamma^4} \qquad \ldots\ldots(10)$$

where η_γ is the viscosity at the velocity gradient γ. The absolute viscosity can thus be found from the equation

$$\eta_\gamma = \frac{h}{2\pi a^4}\left(\frac{dC'}{d\omega} + \frac{3C'}{\omega}\right). \qquad \ldots\ldots(11)$$

§ 5. THE MEASUREMENT OF THE ABSOLUTE VISCOSITY OF THIXOTROPIC MATERIALS

Since the absolute viscosity of an anomalous fluid can be measured comparatively easily if the viscosity is a function only of the applied shearing stress, thixotropic viscosity changes can be followed in a thixotropic sample provided the viscosity of the sample *at every point in the viscometer* is controlled so as to be a function of the applied shearing stress alone.

The concentric cylinder viscometer allows this condition to be attained fairly readily. If a given constant couple C per unit length is applied to the specimen for a long period the effect of previous shear history can be wiped out, and an equilibrium viscosity for the appropriate shearing stress f_0 can be obtained for all values of r between a and b.

If after this conditioning of the sample, the applied couple is altered suddenly to a fraction $1/m$ of its previous value, the applied shearing stress f at any radius r will be $1/m$ of its previous value.

But eqn. (1) can be rewritten for these conditions as $\eta = F(mf, f, t)$ which expresses the viscosity relationship for all points in the viscometer. If the time t is also given an arbitrary value the viscosity of the fluid at all points in the viscometer can be represented as a function of the applied shearing stress only, and it can be measured as for an instantaneously thixotropic sample.

In this way it becomes possible, in general, to find not only the equilibrium viscosity at a given shearing stress f_0, but also the viscosity corresponding to the shearing stress f applied for a time t after equilibrium at f_0 had previously been attained, and the relationship $\eta = F(f_0, f, t)$ can be investigated for all values of f_0, f and t.

In a similar manner the parallel disc viscometer can be used to condition the sample to an equilibrium viscosity corresponding to a maintained constant velocity gradient $g_0 = r\omega_0/h$ at a radial distance r if ω_0 is the relative angular velocity between the discs and is maintained constant. If the angular velocity is suddenly changed to a new and constant value ω, the velocity gradient g at any point will be altered in the ratio ω/ω_0, and if the time t for which this new velocity gradient is maintained is given an arbitrary value the viscosity relationship of eqn. (2) is simplified and the viscosity becomes a function of the velocity gradient g only and can be measured.

Using these techniques it is possible to study the relationships given by eqns. (1) and (2) and to investigate the relaxation processes which take place in thixotropic 'build up' and 'breakdown'.

REFERENCES

Clark, O. H., and Deutsch, M. L., 1950, *J. Appl. Phys.*, **21,** 713.

Eisenschitz, R., 1933, *Kollzeitschr.*, **64,** 184.

Krieger, I. M., and Maron, S. H., 1952, *J. Appl. Phys.*, **23,** 147.

Lindquist, C. G., and Sierichs, W. C., 1950, *J. Coll. Sci.*, **6,** 33.

Mooney, M., 1931, *J. Rheology*, **2,** 210.

Oldroyd, J. G., 1949, *J. Coll. Sci.*, **4,** 333.

Rabinowitsch, B., 1929, *Z. Phys. Chem.*, **145A,** 1.

Reiner, M., 1930, *J. Rheology*, **1,** 250.

Richardson, E. G., and Tyler, E., 1933, *Proc. Phys. Soc.*, **45,** 142.

Schofield, R. K., and Scott Blair, G. W., 1930, *J. Phys. Chem.*, **34,** 248.

Søltoft, P., 1948, Proc. International Rheological Congress Scheveningen (Holland) 1948, Part II, 273.

Thornton, S., and Rae, D., 1953, *Proc. Phys. Soc.* B, **66,** 120.

The Measurement of the Absolute Viscosity of Anomalous Fluids

II: A Comparison of Absolute Viscosity Measurements on Instantaneously Thixotropic Fluids

BY S. THORNTON AND D. RAE
Department of Physics, Durham Colleges in the University of Durham

Communicated by W. A. Prowse; MS. received 15th August 1952

ABSTRACT. The assumptions inherent in the Rabinowitsch and Mooney methods of measuring the absolute viscosity of anomalous fluids are considered critically, and practical methods are suggested for overcoming the difficulties which arise in these determinations. Absolute viscosity measurements have been obtained on instantaneously thixotropic samples with a suitably modified Poiseuille tube apparatus and a concentric cylinder viscometer, and the results are shown to be in agreement within the limits of experimental error over the range of shearing stresses considered.

§ 1. INTRODUCTION

BEFORE attempting a serious investigation of the phenomenon of thixotropy by studying the time dependence of the viscosity of thiotropic fluids as outlined in Part I (Thornton 1953, to be referred to as I), it was considered necessary to check the validity of the method. Since a suitable way of doing this was by showing that the absolute viscosity of instantaneously thixotropic fluids is independent of the apparatus used for its measurement, experiments were designed to check whether the viscosities of such materials, when measured with a concentric cylinder viscometer, were in agreement with those obtained with a Poiseuille tube apparatus. This investigation necessitated the preparation of suitable samples and the study of their behaviour under the special conditions of the experiment.

§ 2. THE CHARACTERISTICS OF INSTANTANEOUSLY THIXOTROPIC FLUIDS

The phenomenon of instantaneous thixotropy can be recognized in many dispersions of finely divided solid particles in liquid media. It is undoubtedly due in part to the existence of strong inter-particle attractions, but it is likely that the rigidity of the solid particles also plays an important part, since emulsions and dispersions of colloidal macromolecular materials which lack this rigidity in the dispersed phase tend to exhibit visco-elasticity and normal rather than instantaneous thixotropy, peculiarities which render these materials unsuitable for investigation by means of the Poiseuille tube apparatus.

Strong inter-particle attractions in a solid–liquid dispersion give rise, in the absence of applied shearing stress, to flocculation. If the concentration of the dispersed phase is low this leads to sedimentation of the solid phase, giving rise to a heterogeneity in the sample on a macroscopic as well as on a microscopic scale. If the concentration is high the flocculates may be so closely packed together that they form a connected structure throughout the volume of the dispersion. Such a dispersion is stable and acts as a homogeneous fluid on a macroscopic scale. The low concentration suspensions are unsuitable for prolonged viscosity experiments because of the complications arising from sedimentation. The high concentration suspensions on the other hand are suitable, but the high concentration of the dispersed phase automatically introduces complications because at low shearing stresses an anomalous fluid of this type behaves as a rigid structure or *false body*, and it appears to be a general characteristic of instantaneously thixotropic materials that they have a double yield value: a static yield value f_s defined as the shearing stress required to start flow in the material at rest, and a dynamic yield value f_d defined as the shearing stress at which flow ceases in the moving material. Moreover since solid bodies do not transmit hydrostatic pressures in the same way as a liquid, the behaviour of instantaneously thixotropic samples in Poiseuille tubes, where plug flow must inevitably take place, must be investigated carefully, and it becomes necessary, therefore, to consider critically the assumptions of the elementary Rabinowitsch theory (Rabinowitsch 1929) for the special case of instantaneously thixotropic materials.

§ 3. POISEUILLE-TUBE INSTRUMENT

(i) *Assumptions which govern Instrument Design*

In deriving the expression for the shearing stress at a given radial distance r from the axis it is assumed in the Rabinowitsch treatment that the pressure over any plane perpendicular to the axis of the tube is constant, but because the solid plug of an instantaneously thixotropic material may affect the distribution of the hydrostatic pressure forcing the fluid along the tube, this condition is normally invalid. In fact the driving force on the solid plug can be regarded in part as a direct (solid) thrust on it by the material in the reservoir supplying the tube. Though the magnitude of this thrust is not easily derived mathematically, experiment has shown that it gives rise to errors of 10–15% in the estimation of the pressure gradient required to produce a given efflux rate, and leads to serious errors in the measurement of absolute viscosity.

In order to remedy the improper pressure distribution due to plug flow it is necessary to eliminate the direct thrust on the plug from the reservoir. This can be done by bringing the sample in the tube to the fluid state over the whole cross-sectional area of the tube, at some point along the tube, so that the hydrostatic pressure acting at that point is wholly responsible for forcing the fluid (and the solid plug which forms later beyond this point) along the tube. For this special case the pressure gradient along the remainder of the tube, which is defined here as the effective part of the tube, is uniform and the equations developed by Rabinowitsch and by Clark and Deutsch (1950) are valid.

Experimentally it is found that most instantaneously thixotropic fluids tend to slip at a smooth wall when a shearing stress is applied. A careful study of this phenomenon, using the tube apparatus and also a special apparatus designed for the purpose of studying slip and yield value, seems to indicate that slip is produced by a local breakdown in the homogeneity of the sample, whereby liquid exuded from the bulk material acts as a lubricant. Since it is difficult to obtain consistent results when slip is allowed it is clearly advisable to aim at eliminating it, rather than attempting to measure it as Mooney (1931) did. With the subsidiary apparatus it was found that consistent results could be obtained and there was no evidence of slip, provided the surfaces were roughened to such an extent that the surface irregularities were small on a macroscopic scale but large in comparison with the size of the particles in the suspension used as the anomalous fluid.

(ii) *Assumptions which govern Measuring Technique*

Owing to the existence of a double yield value it is found that the relationship between the velocity gradient g and the shearing stress f is not single-valued for all values of f for instantaneously thixotropic samples: the theory requires that this relationship shall be single-valued. Figure 6 shows the relationships which have been obtained experimentally for two such samples. They are typical of substances of this class, and one of them (sample A) indicates that the material *in the fluid condition* exhibits a single-valued functional relationship between g and f which does not apply to the unstable transition from the false body to the fluid state at the shearing force f_s. Once the material has been brought into the fluid state by the application of a shearing stress greater than f_s it remains in that condition for all values of the shearing stress greater than f_d. Thus for shearing stresses between f_d and f_s the behaviour of the sample depends on its previous state: if it were previously in the fluid state it gives rise to a velocity gradient appropriate to the applied shearing stress, but if in the false-body state it does not yield. In an instrument such as the tube apparatus where both fluid and false-body states occur together, difficulties may arise if the conditions are such that material in the false-body state is being broken down into the fluid state during the experiment; for example, immediately after the rate of flow of the substance through the tube has been increased, and before the equilibrium conditions governing the radius of the solid plug have been established along the full length of the tube. If sufficient time is allowed for equilibrium conditions corresponding to a steady efflux rate to be attained no difficulty arises, and since it is considered that all the material entering the effective part of the tube does so in the fluid state, the solid plug must be formed from fluid material, and the conditions in the tube are those appropriate to a single-valued functional relationship between g and f. Reducing the shearing stress in the tube does not invalidate this condition, and since the adjustment in the radius of the solid plug can be expected to take place along the full length of the tube immediately the pressure gradient is reduced, this eliminates the period of waiting to ensure equilibrium conditions. This method was used in practice.

§ 4. CONCENTRIC CYLINDER INSTRUMENT

(i) *Assumptions which govern Viscometer Design*

The instrument is so designed that the outer cylinder, which contains the sample, rotates about a vertical axis, and that the inner cylinder which is suspended in it can be aligned on the same axis. The region round the lower ends of the two cylinders gives rise to an end effect in the treatment of the flow between concentric cylinders, but this can be made comparatively small by using long cylinders, and the effect can be measured and eliminated by the method of differences if readings are taken for different depths of fluid. Experimentally it was found that the end correction was equivalent to an increase in the depth of sample used, and that this correction factor wa sindependent of the speed and of the quantity of material used, within wide limits, for any particular sample.

The consideration of slip at the bounding cylinders is the same as that for the tube wall, and if the surfaces are suitably roughened slip can be eliminated.

(ii) *Assumptions which govern Measuring Technique*

As in the case of the tube instrument the unwanted values of the (velocity gradient, shearing stress) curve, which arise only when material in the false-body state is being broken down into the fluid state, can be eliminated by using a suitable measuring technique. When the radius of the boundary between the fluid and the false-body state is being increased, the example of sample A in fig. 6 indicates that the shearing stress at the boundary is the static yield value; when it is being decreased the stress is the dynamic yield value. It can be shown that for any shearing stress at the inner cylinder the relative angular velocities of the two cylinders when the volume of the material in the false-body state is (*a*) increasing, and (*b*) decreasing, differ by a constant amount. Thus since the absolute viscosity is determined from the slope of the curve relating applied couple and angular velocity in this viscometer, it can be obtained by utilizing the conditions corresponding to either (*a*) or (*b*) above.

For the first condition to hold the sequence of events must be such that although readings are taken at constant angular velocity of the outer cylinder, the progressive changes in this velocity must represent a decrease, whereas for the second condition there must be a progressive increase.

Consider now progressive changes involving an increase in the angular velocity. Starting at the lowest shearing rate a difficulty arises because all the material in the viscometer is originally in the false-body state and before any of it can be obtained in the fluid state, a shearing stress f_s must be applied at the inner cylinder. But the fluid material obtained at this shearing stress is only in equilibrium at the appropriate velocity gradient which is not necessarily the arbitrary velocity gradient which is produced at the moment of breakdown by the rotation of the viscometer cup. Readjustment of equilibrium is brought about by a return swing of the inner cylinder, but this also introduces an uncertainty in the value of the shearing stress applied at the interface between the fluid and false-body states. A certain increase in the angular velocity is then required before it can be assumed that the shearing stress at the false-body surface is the static yield value.

Similarly when the progressive changes in the angular velocity of the outer cylinder represent a decrease difficulties are encountered at the highest velocities. The highest angular velocity produces a shearing stress f_s in the material at the false-body surface, and in the fluid bounding this surface the velocity gradient is that appropriate to this shearing force; but as the angular velocity of the outer cylinder is reduced, the shearing stress at the false-body surface is also reduced, and although the fluid state is not transformed into the false-body state as long as f is greater than f_d nevertheless the velocity gradient at the false-body surface is altered. Thus it is not until the angular velocity of the outer cylinder has been reduced sufficiently for the position of the interface between the fluid and the false-body states to move that reliable results, consistent with the single-valued function relating g and f for the fluid state, can be obtained.

Provided these difficulties are avoided either the progressively increasing or the progressively decreasing angular velocity method is suitable for measuring absolute viscosity, and in fact both methods have been employed.

§ 5. THE DESIGN OF THE TUBE INSTRUMENT

The essentials of the tube instrument are shown in fig. 1. The flow tube A is of glass and is connected directly to the reservoir B along which a close fitting plunger C can be driven at any of a set of predetermined rates governed by a constant speed motor, to give definite known rates of flow in the flow tube. In order that the equations of viscous flow developed by Rabinowitsch may be valid it can be shown that the samples must be brought to the fluid state at all radii at some axial distance along the flow tube. A grid or an obstacle E is therefore inserted near the reservoir end of the tube to restrict the cross-sectional area of the tube sufficiently to produce fluid flow, as distinct from plug or false-body flow, immediately beyond it. The portion beyond this point is the effective part of the flow tube, and to prevent slip it is necessary for its inner surface to be roughened.

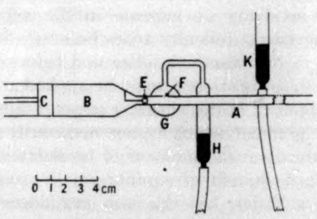


Fig. 1. The tube instrument.

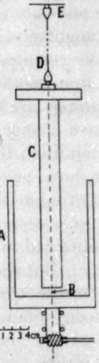


Fig. 2. The concentric cylinder instrument.

The excess pressure in the material undergoing flow at a distance *D* from the end of the tube open to the atmosphere is equal to the pressure exerted by the wall of the tube on the material at this axial distance. Hence, if a small portion of the tube wall is removed at this point, and an external air pressure equal to the pressure in the material applied at the hole so made, the flow will not be disturbed. The pressure in the material, due to the flow, can then be found by measuring the applied pressure. In the instrument a hole F is bored in the wall of the tube near the reservoir end, but in the effective part of the tube. This hole and a short portion of the tube containing it are surrounded by a small closed pressure chamber G which is connected to a variable manometer H, K, which enables the applied pressure to be adjusted.

The experimental procedure is to set the rate of flow at its maximum value and to adjust the pressure in the chamber, by means of the movable limb K of the variable manometer, until the boundary between the material in the tube and the air in the chamber is stationary and at the position formerly occupied by the tube wall at the hole. This pressure is measured on a fixed manometer not shown in the diagram. The next lower value of the rate of flow is then selected and the pressure in the chamber readjusted for equilibrium at the hole for this new rate. Provided the whole apparatus is suitably temperature controlled a series of values of the rate of flow and the corresponding pressures can be obtained enabling the relation between the viscosity and the shearing stress at a definite temperature to be determined.

§ 6. THE DESIGN OF THE CONCENTRIC CYLINDER VISCOMETER

The essentials of the concentric cylinder viscometer are shown in fig. 2. The outer cylinder A, which contains the sample and rotates about an axis B which can be adjusted to be vertical, is driven by a special motor which acts essentially as a constant speed motor with an infinitely variable gear ratio over a ratio of speeds 400:1. The inner cylinder C, which is roughened sufficiently to prevent slip at the walls, is suspended by a long torsion wire D from a rigid support E which can be adjusted so that the axis of this cylinder is coaxial with that of the outer cylinder. The angular deflection of the suspension wire is used in the usual way to determine the torsional couple per unit depth of sample at any angular velocity of the outer cylinder.

Though the ratio of the radii of the inner and outer cylinders is not critical, this was taken as $1:\sqrt{10}$ to simplify the correction for changing the measured angular velocities to those for a semi-infinite sea viscometer, but in practice it was found that the correction terms for the instantaneously thixotropic samples investigated were all negligible, the concentric cylinder apparatus itself acting as a semi-infinite sea viscometer.

To ensure accurate control of the temperature of the specimens, all of which exhibited a large temperature dependence of absolute viscosity and had a low thermal conductivity, it was necessary to keep the temperature of the inner cylinder as well as the outer constant. This was done in practice by surrounding the outer cylinder with a constant temperature oil bath maintained at the same temperature as a lagged enclosure which surrounded the upper part of the inner cylinder and the material under test, and which had its own auxiliary temperature controlling apparatus. With this arrangement the temperature of the system was controlled to $\pm 0{\cdot}1^\circ$ C, and the viscosity variations of the samples on this account were estimated, from experiments performed over a range of temperatures, at 0·75% at 25° C.

§ 7. EXPERIMENTS AND RESULTS

Since with a particular instrument the accuracy of measuring the absolute viscosity of an instantaneously thixotropic fluid at a given shearing stress depends on the viscosity of the sample, as indeed it does also when Newtonian fluids are under investigation, it is evident that the accuracy for such a material will vary with the shearing stress employed. Because of this, care is necessary in matching an instrument to the rheological properties of the sample under investigation if accurate results are to be obtained over the whole range of shearing stress applied. In these experiments a number of pigment-in-oil dispersions were studied. They all exhibited the phenomenon of instantaneous thixotropy, and were suitable for examination, but eventually the criterion of accuracy over the whole range of shearing stress led to a special study of samples with viscosities ranging from infinity to values of the order of 10–20 poise, and with yield values of the order of 100 dyn/cm^2; such samples did not sediment out and allowed the absolute viscosity over the range of shearing stresses used in the concentric cylinder and flow tube apparatus to be measured with an experimental error in each case of about 2 or 3% under the worst conditions.

Difficulties were encountered with recently prepared dispersions because of ageing, a process which resulted in progressive changes in the viscosity–shearing stress relationship with time, and which is usually attributed to chemical reactivity between the pigment and the suspension medium. To minimize this reactivity samples of titanium dioxide pigment in liquid paraffin were prepared. Control of the rheological properties to match the requirements of the viscometers was obtained from the choice of pigment grade, the pigmentation ratio, the viscosity of the suspension medium, and by the addition of small amounts of wetting agent, and suitable samples were prepared for accurate comparisons of the absolute values of the viscosity over a range of shearing stresses wide enough to cover the region where most of the change of viscosity with shearing stress takes place.

In the early experiments where the grid E (fig. 1) was omitted, reproducible results were obtained for the two instruments separately, but though both sets indicated the same general kind of relationship between shearing stress and viscosity, the graphs appeared to be displaced considerably and it was doubtful which set of results was the more reliable. Eventually it was suspected that the flow-tube apparatus was at fault and experiments were carried out to test the effect of inserting a grid in the tube. It was found that the results for the tube instrument had been in error and the insertion of the grid brought them into agreement with those for the concentric cylinder apparatus. The configuration of the grid or its position in the tube in front of the pressure measuring device was not important.

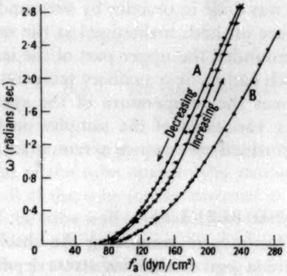


Fig. 3.

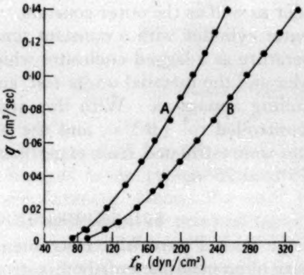


Fig. 4.

The comparison of results for the two instruments has been performed carefully a number of times using samples specially prepared to match the requirements of the instruments, and has confirmed the original conclusion that the results obtained with the two instruments are in agreement. Two such comparisons are given. Figure 3 shows the relationships between ω and f_a obtained for two samples A and B with the concentric cylinder viscometer; and for sample A values of ω for the progressively increasing and progressively decreasing conditions are given to show the hysteresis obtained in a cycle due to the double yield value. This hysteresis is reproducible and is not to be confused with the hysteresis effect which is not readily reproducible and which arises because of the time dependence of viscosity when thixotropic samples are used. The relation between ω and f_a for sample B (which differs from sample A in concentration of pigment, condition of milling, and time of ageing) is for the special case where only the dynamic yield value is operative. Figure 4 gives the relations between the efflux rate q and the shearing stress at the wall of the flow tube f_R for the same samples when precautions are taken once again to ensure that only the dynamic yield value is effective. Figure 5 shows the relations between the shearing stress and the absolute viscosity of the two samples when the latter is determined (*a*) from the curves of fig. 3 relating ω and f_a, and denoted by η_α and (*b*) from the curves of fig. 4 relating q and f_R and denoted by η_β; η_α is determined by the method outlined in I and η_β is given by an extension of the Rabinowitsch treatment as

$$\eta_\beta = \pi R^3 \left(\frac{dq}{df_R} + \frac{3q}{f_R}\right),$$

where R is the radius of the flow-tube.

The percentage differences between η_α and η_β are tabulated in the table. They are of the order of the experimental error in the determination in both cases.

Percentage Difference between η_α and η_β at Different Shearing Stresses for Samples A and B

f	80	100	120	140	160	180	200	220	240	260	280
Sample A	−1	−6	1	1	0	4	−1	−5			
Sample B		−2	6	−4	1	1	0	0	2	1	1

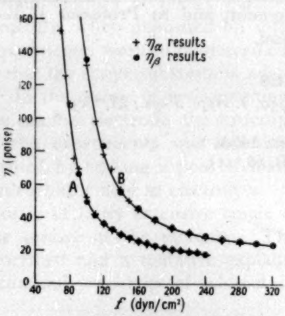


Fig. 5. Variation of the absolute viscosity with shearing stress for samples A and B showing the results obtained: η_α with the concentric cylinder instrument, and η_β with the tube instrument.

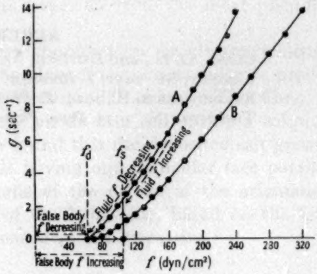


Fig. 6.

Since the results obtained are typical of instantaneously thixotropic fluids as a class and since the definition of the viscosity of anomalous fluids is somewhat arbitrary, the more fundamental relationship between velocity gradient and shearing stress is shown in fig. 6 to illustrate the behaviour of such materials.

For the sake of completeness the yield values f_d and f_s of sample A are also given. f is readily calculated from the hysteresis loop in fig. 3 since the difference in the value of w at a given shearing stress for the progressively increasing and progressively decreasing conditions can be shown to be a constant, in this case 0·25 radian/second. This value of ω on the progressively decreasing curve corresponds to the shearing stress 106 dyn/cm^2, which is f_s the static yield value. The dynamic yield value is obtained most easily by plotting fluidity (defined as the reciprocal of the viscosity) against shearing stress and extrapolating the curve to the f axis, which it appears to cut at a large angle. For sample A this value is 62 ± 2 dyn/cm^2 at $25 \pm 0{\cdot}1°$ c, the temperature of the experiment.

§ 8. CONCLUSION

The measurement of the absolute viscosity of anomalous fluids by means of a Poiseuille tube apparatus is more complicated than would appear at first sight, but the difficulties which arise do so for the most part because of the peculiar behaviour of instantaneously thixotropic materials, the only class of anomalous fluids which can be examined in this way. The concentric cylinder viscometer on the other hand affords a comparatively easy way of measuring absolute viscosities and since the results obtained with such an instrument have been shown to be absolute and independent of the apparatus used a method is now available as suggested in I for the quantitative study of the time dependence of the viscosity of thixotropic materials.

ACKNOWLEDGMENTS

We wish to express our thanks to the Council of the Durham Colleges for a grant in aid of this research, to the British Titan Products Company Limited for gifts of samples of titanium dioxide pigment, and to Professor J. E. P. Wagstaff for the facilities placed at our disposal.

REFERENCES

Clark, O. H., and Deutsch, M. L., 1950, *J. Appl. Phys.*, **21**, 713.
Mooney, M., 1931, *J. Rheology*, **2**, 210.
Rabinowitsch, B., 1929, *Z. Phys. Chem.*, **145A**, 1.
Thornton, S., 1953, *Proc. Phys. Soc.* B, **66**, 115.

The Structure and Orientation of Calomel formed on Liquid Mercury by Anodic Polarization

By H. R. THIRSK

Laboratories of Physical Chemistry, King's College, University of Durham,
Newcastle-on-Tyne

MS. received 10*th July* 1952, *and in final form* 7*th October* 1952

ABSTRACT. The electrochemical formation of mercurous chloride on mercury in the presence of dilute hydrochloric acid as electrolyte has been found to give rise to a surface orientation in which the a_0b_0 face diagonal and the c_0 axis of the tetragonal unit cell lie in the plane of the mercury. Fairly extensive single crystals may be formed or a more polycrystalline product may constitute the film, but in either case only {110} faces are found parallel to the substrate. Examples of twinning on {112} planes and rotational slip on {110} planes are noted in the growth of the film. The crystal orientation appears to be due to an approximately hexagonal arrangement of pairs of mercury and chlorine atoms in {110} faces of the mercurous salt. Coherence of the film is assisted by the ready twinning and the presence of the {110} and {011} cleavage planes of the calomel.

§ 1. INTRODUCTION

Examples of the growth of specifically oriented deposits of substances on solid single crystal substrates by electrodeposition, evaporation or chemical reaction are well known and numerous studies have been reported. In addition some substances show preferred crystal habits with certain families of planes parallel to the interface, but otherwise random in structure, when deposited on solid substrates such as glass or collodion. The experimental work to be described differs somewhat from the above phenomena in that the active substrate is a liquid.

In the course of investigations in this laboratory on the electrochemistry of the calomel electrode, the structure of the films of mercurous chloride produced in the experiments was examined by electron diffraction. The films were formed by making a pool of mercury the anode in a cell having a solution of hydrochloric acid as electrolyte. It was found that the deposited salt grew as a mosaic of fairly extensive single crystals having one particular face parallel to the surface of the mercury. The details of the nature of the orientation is described and a tentative explanation of the behaviour, based on the known structural properties of calomel, is advanced in the discussion.

§ 2. EXPERIMENTAL

The films of calomel were prepared by using an H-shaped glass vessel in which cylindrical anode and cathode compartments were separated by a connecting tube. The larger compartment, 3 cm in diameter, contained a pool of mercury functioning as an anode to which electrical contact was made by means of a sealed-in platinum wire; the cathode compartment contained a large platinum electrode. The mercury was purified first by dropping it, as a fine spray, through nitric acid and then by distilling it in glass at a pressure of a few cm Hg. It was, therefore, reasonably free from grease and foreign ion impurities. The electrolyte was prepared by diluting distilled constant boiling hydrochloric acid with conductivity water to produce a 2N solution. The mercury was polarized at a current density of 2·5 mA/cm^2, the polarization being continued sufficiently long to produce layers of from 60 to 200 Å in thickness as calculated from Faraday's laws. The process could be watched by noting changes in the uniform interference colours as the film thickened.

After the layers had been formed they were removed from the mercury surface by gently lifting on a platinum gauze made from fine wire. They were then floated off the gauze on to a clean water surface in order to wash the film, and picked up on a microscope slide where portions holding excess of mercury were removed by means of a dissecting needle. Supported in this manner they were dried by leaving in a continuously evacuated desiccator over phosphorus pentoxide for several hours. Finally the films were coated with a 1% collodion solution and stripped off on to the E.M.3 type copper grid specimen support by the Scotch tape method. The grid was separated from the calomel by a collodion layer so that chemical interaction was impossible.

The specimens were examined by electron diffraction and microscopy in a Metropolitan-Vickers E.M.3 electron microscope at 75 kv accelerating potential and by diffraction in a Finch-type electron-diffraction camera worked at 50–55 kv. Figures 1 to 6 (Plate) show diffraction patterns obtained from the films for which, in every case, the electron beam was essentially perpendicular to the plane of the specimen. Figures 1 and 5 also show a calibrating thallous chloride ring pattern, the thallous salt being deposited on the film from a very dilute aqueous solution. The main features of the structures of the deposits are described by reference to these figures.

§ 3. DISCUSSION

The structure of mercurous chloride has been examined by X-ray diffraction by several authors, Hylleraas (1926) and Havighurst (1925, 1926). The substance has a tetragonal unit cell, space group D_{4h}^{17} with $a_0 = 4{\cdot}47$ Å and $c_0 = 10{\cdot}89$ Å. The diffraction pattern reproduced in fig. 1, may be readily indexed by means of fig. 9. It corresponds to the expected pattern produced by an electron beam perpendicular to a (110) face of a single crystal corresponding to the above description. Thus the *H* direction of fig. 9 corresponds to the $[1\bar{1}0]$ lattice row and the *L* direction to the [001] lattice row of the reciprocal lattice, (*hh*0) and (00*l*) even. The dimensions of the unit cell calculated by comparison with the simple-cubic thallous chloride pattern ($a_0 = 3{\cdot}834$ Å) are $a_0 = 4{\cdot}47$ Å, $c_0 = 10{\cdot}86$ Å, which is in very close agreement with the X-ray data.

The largest crystals found in the deposits were about 0·2–0·3 mm across. Some specimens did not consist extensively of crystals having an area sufficiently great for patterns such as fig. 1 to be obtained, considerable areas being composed of a mosaic of smaller fragments. A diffraction pattern from one of these more finely crystalline areas is reproduced in fig. 2. It is easily shown that no planes other than the {110} are parallel to the surface, for fig. 2 can be obtained quite simply from fig. 1 by the rotation of the rectangular spot pattern about an axis parallel to the electron beam and passing through the centre of the pattern.

The properties of large crystals of calomel are quite well known. According to Dana (1951) the material has a good {110} cleavage, a rather poor {011} cleavage and twins on the {112} face. The twinning occurs either as contact or interpenetration twins, often repeated, and with irregular, often concealed, boundaries. The crystals are stated to be plastic. It is thus seen that the face parallel to the mercury surface is the important cleavage face of the crystal.

Figure 7 has been drawn to represent diagrammatically the arrangement of the mercury and chlorine atoms in this face; the atoms are situated in pairs. The convention used in the drawing shows the positions of the centres of the atoms by means of the crosses, and represents the pairs of atoms by circles drawn through the crosses as continuous lines for the mercury and dashed lines for the chlorine atoms. The Hg–Hg distance is 2·53 Å, the Cl–Cl distance 3·33 Å and the Cl–Hg distance 2·52 Å. It is seen from the drawing that due to the fact that d_{110} and d_{001} are very nearly in the ratio of $1:\sqrt{3}$ an approximately hexagonal arrangement is taken up by the mercury and chlorine ion pairs in the {110} faces. The traces of two {112} twinning planes perpendicular to this face are parallel to the diagonals of the rectangular lattice net, lying, therefore, at an angle of 59° 49′ to the a_0b_0 face-diagonal lattice row.

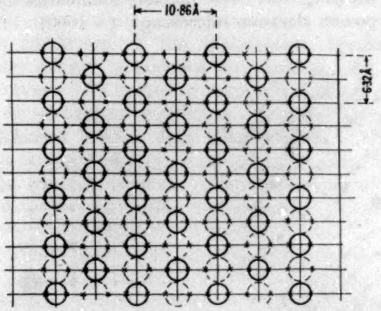


Fig. 7. Diagram of the arrangement of the Hg and Cl atoms in the {110} cleavage face of Hg_2Cl_2.
Continuous circles drawn through the centre of the pairs of Hg atoms. Position of centres marked with a cross.
Dashed circles drawn through the centres of the pairs of Cl atoms. Position of centres marked with a cross.
Twinning planes are found perpendicular to this face and parallel to the diagonals of the rectangular lattice.

It would be expected that evidence for doublets and triplets with 59° 49′ between the c_0 axis should be found on examining the calomel films. Figure 4 is the reproduction of a pattern from a triplet of this type with an almost hexagonal pattern resulting from the superposition of three patterns similar to fig. 1, relatively at 59° 49′.

Wilman (1950, 1951) and Evans and Wilman (1950) have shown that strain is often relieved during the formation of oriented overgrowths by the rotational slip of portions of the crystal. This occurs on certain faces when alternative positions of metastable equilibria are taken up and, in some cases, continuous rotation between these metastable positions is produced. It seems very likely that a behaviour of this kind would occur in the growth of the mercurous chloride films, because of the development of the crystal on a cleavage face and it is considered that figs. 5 and 6 provide evidence for this. These diffraction patterns were taken from portions of the specimen which, by means of an examination of low power electron microscopy, appeared to be free from sharply defined crystal boundaries thus distinguishing the specimens from, for example, the one giving rise to fig. 3. Figure 5 is a pattern from a crystal having a portion rotated at an angle, using Wilman's notation, of $\delta = 3° 20'$ about an axis perpendicularly to the (110) interfacial plane. δ is defined for an atomic lattice with rectagular axes, a, b, corresponding in this case to the $\{1\bar{1}0\}$ and c_0 lattice rows in the (110) plane, in such a way that, if the two superimposed sheets are rotated in opposite directions by $\frac{1}{2}\delta$ from the parallel azimuth so that the $[u_1v_1]$ and $[u_1\bar{v}_1]$ rows coincide along the initial a axis direction, $\tan\frac{1}{2}\delta = v_1b/u_1a$. Similarly, for this same rotation, two rows $[u_2v_2]$ and $[\bar{u}_2v_2]$ will coincide along the b axis and $\tan\frac{1}{2}\delta = u_2a/v_2b$. This gives $\delta = 3^\circ\ 20'$, $u_1/v_1 = 60$; $v_2/u_2 = 20$.

Figure 3 was taken at the junction of two crystals in the film and is due to the joining of the crystals with the respective c_0 axes at an angle of 72°. The pattern has been indexed in fig. 8 on this assumption by the appropriate superposition of the reciprocal lattice nets corresponding to the single-crystal pattern of fig. 1. Patterns similar to fig. 3 were frequently observed and corresponded to portions of the specimens showing very well marked boundaries between crystals as observed by low power electron microscopy ($\times 1000$). Let the coinciding lattice rows in the two lattices again be $[u_1v_1]$, $[u_1\bar{v}_1]$, and $[u_2v_2]$, $[\bar{u}_2v_2]$ where $\tan 36^\circ = v_1b/u_1a$ or u_2a/v_2b, the lattice rows c and $[1\bar{1}0]$ in the $\{110\}$ plane again being denoted by b and a. Thus for this case $v_1/u_1 = 0{\cdot}422 \simeq 3/7$; $u_2/v_2 = 1{\cdot}250 = 5/4$ and therefore the coincident lattice rows are of relatively low indices and identity translation corresponding to a relatively stable intercrystalline boundary. In order to show the indexing of the diffraction spots in fig. 6 and for increased clarity, the pattern is drawn diagrammatically in fig. 9, suggesting positions of metastable equilibria during rotation as follows, the values being probably less accurate than for figs. 3, 4 and 5: $\delta' = 16^\circ\ 20'$ corresponding to $u_1/v_1 = 12$, $v_2/u_2 = 4{\cdot}05 \simeq 4$; $\delta'' = 22^\circ$ corresponding to $u_1/v_1 = 8{\cdot}85$, $v_2/u_2 = 3$; and $\delta''' = 13^\circ\ 10'$ corresponding to $u_1/v_1 = 14{\cdot}95 \simeq 15$, $v_2/u_2 = 5{\cdot}04 \simeq 5$.

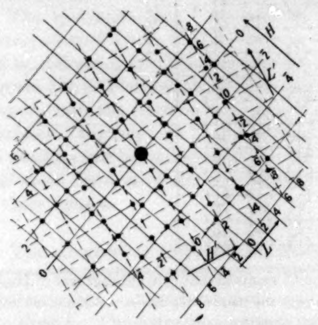


Fig. 8. Indexing of fig. 3 consisting of two rectangular spot patterns as shown in fig. 9 superimposed with the H direction of one at 72° to the similar direction of the second pattern marked as H'.

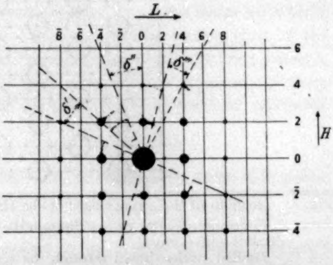


Fig. 9. Drawing of fig. 6 to show more clearly the angles δ', δ'' and δ''' between positions of metastable equilibrium during rotation. H corresponds to $[hho]$, h even, L to $[00l]$, l even.

Calomel readily voltalizes and Kirchner (1932) has studied the orientation of the substance on a collodion substrate. He found that the deposit had a one-dimensional orientation with the c_0 axis lying in the plane of the substrate. A similar behaviour would not have been surprising in the experiments described above, but the further restriction of the a_0b_0 face diagonal also lying in the plane of the interface is particularly interesting since it determines uniquely the interfacial plane and manner of growth perpendicular to the surface.

It would seem that the fortuitous symmetry of the {110} face brings about a peculiarly favourable set of conditions for the formation of this plane during the electrolysis of the underlying mercury, particularly if certain assumptions are made concerning the nature of the surface of the liquid mercury. These assumptions are necessary because, although a study of the surface of mercury has been attempted by several groups of workers, unfortunately no conclusive opinions have been reached. Bailey, Fordham and Tyson (1938) make a summary of previous work by both x-ray and electron diffraction. They suggest that normally the surface has a skin of surface impurity; underneath the skin the metal surface is truly amorphous. If this is so, the following processes may occur in the present experiments. On polarization the surface impurity, if existing under the conditions of the experiments, breaks down and the close packed mercury surface is exposed for reaction. By x-ray diffraction, Campbell and Hildebrand (1943) showed that each mercury atom in the liquid is 3·0 Å away from its six nearest neighbours. A similar arrangement of atoms at the surface could react, with a slight displacement, at certain places with chloride ions to give sheets of mercurous chloride with close fitting between the mercury atoms in the calomel and in the surface of the mercury. These ordered layers may then form centres for further crystal growth, increasing in area rather than depth during the initial stages of the process. Some relative movement to assist the orientation between separate growing units may take place by rotation but, in addition, the crystals of calomel possess cleavage and twinning properties described above assisting in the formation of a coherent film. Thus rotation on {110} faces has been demonstrated and there seem to be several metastable positions which crystal units may take up by rotation within the crystal on planes parallel to the interface; twinning occurs on {112} faces perpendicular to the mercury; finally strain may further be relieved at an angle to the face in contact with the mercury surface by twinning on {112} planes at 30° 11′ to the surface, by slip on the {011} cleavage faces at 45° to the surface, and by slip on the {110} faces perpendicular to the surface.

With the exception of the corresponding salts of bromine and iodine, it would seem rather unlikely that further equally well defined orientations of overgrowths on liquid mercury are to be found unless investigation shows that favourable structural details exist in other slightly soluble mercury compounds. Mercurous sulphate, for example, with a monoclinic unit cell grows on the surface of mercury under similar conditions as fine, discrete crystals in completely random array. The significance of the purely electrochemical considerations associated with this work will be discussed elsewhere.

REFERENCES

BAILEY, G. L. J., FORDHAM, S., and TYSON, J. T., 1938, *Proc. Phys. Soc.*, **50,** 63.
CAMPBELL, J. A., and HILDEBRAND, J. H., 1943, *J. Chem. Phys.*, **11,** 330.
DANA, E. S., 1951, *System of Mineralogy* (New York: Wiley), 7th edn., Vol. II, p. 25.
EVANS, D. H., and WILMAN, H., 1950, *Proc. Phys. Soc.* A, **63,** 298.
HAVIGHURST, R. J., 1925, *Amer. J. Sci.*, **10,** 15 ; 1926, *J. Amer. Chem. Soc.*, **18,** 2113.
HYLLERAAS, E., 1926, *Z. Phys.*, **36,** 862.
KIRCHNER, F., 1932, *Z. Phys.*, **76,** 576.
WILMAN, H., 1950, *Nature, Lond.*, **165,** 321 ; 1951, *Proc. Phys. Soc.* A, **64,** 329.

The Optical Constants of Lead Sulphide, Lead Selenide and Lead Telluride in the 0·5–3 μ Region of the Spectrum

By D. G. AVERY
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 7th August 1952

ABSTRACT. Measurements have been made of the refractive and absorption indices *n*, *k* of PbS, PbSe and PbTe at wavelengths between 0·5 and 3 μ, using a reflection method. The values found for *n* at 3 μ are for PbS 4·10, PbSe 4·59 and for PbTe 5·35. The absorption spectra found agree with earlier work, and it has been shown that for PbS a variation of 200 : 1 in the concentration of impurity centres in specimens produced no change in the magnitude of the absorption.

§ 1. INTRODUCTION

RECENT work by Paul, Jones and Jones (1951) and Gibson (1952) on the transmission of thin slices of crystals of PbS, PbSe and PbTe in the infra-red has established the main features of the absorption spectra of these photoconductive materials. The positions of the main absorption edges at about 0·8 μ for PbS, 1·0 μ for PbSe and 1·5 μ for PbTe had been earlier determined by Gibson (1950), and the more recent work has shown the existence of secondary absorption edges at a wavelength for each material corresponding to its photoconductive limit. It was in order to complete the study of the optical properties of this class of photoconductor that the present work was undertaken. The main object was to obtain more accurate information about their refractive indices, and also to cover those regions of high absorption in which the absorption coefficient of thin crystal slices cannot readily be measured. Some earlier work has already been reported (Avery 1951).

Since the optical method adopted involved reflection from the surfaces of crystal specimens, a preliminary investigation was made of the effects introduced by various polishing techniques, in an attempt to provide good plane reflecting surfaces.

For PbS a detailed study has been made for a number of specimens of varying conductivities. A similar detailed study is not presented for PbSe and PbTe, since the wavelength range covered does not include the long wavelength photoconductive limits of these materials, and the same variety of samples is not at present available.

§ 2. EXPERIMENTAL TECHNIQUES

The reflection method employed for determining the optical constants has recently been described (Avery 1952) and will not be further discussed here. It will suffice to state that the wavelength range covered was that of the room temperature PbS photoconductive cell used as a detector, and that the wavelength resolution of the double monochromator employed (Roberts 1952) was better than 5% throughout.

§ 3. EXPERIMENTAL RESULTS

The quantities measured in these experiments were the refractive index *n* and the absorption index *k*. These are related to the complex dielectric constant ϵ by the relation $\sqrt{\epsilon}=n(1-ik)$, and to the absorption coefficient $K\,\text{cm}^{-1}$, at a wavelength λ, by $K=4\pi nk/\lambda$. The results are presented as graphs either of *n* against wavelength, or of the real and imaginary parts of the dielectric constant (in this notation $n^2(1-k^2)$ and $2n^2k$), against wavelength.

(i) *Effect of Surface Preparation*

Since it is not generally possible to obtain very flat large area cleavage faces of crystals of these materials, measurements were first carried out using polished surfaces. It was soon found that the measured values of *n* and *k* depended on the particular surface treatment used. Figure 1 (*a*) shows the values obtained for *n* at various wavelengths for all three substances using two different polishing techniques 'Heavy' and 'Light'. The first consisted of grinding the crystal face flat with the full range of abrasives from 1F to 4F, and then vigorously polishing on a rotating cloth lap using a liquid polish, the second was a slower process, using only 3F and 4F for grinding, and polishing with a light pressure again on a cloth lap using 4F as abrasive. It is seen that the two methods of polishing give significantly different results for each compound.

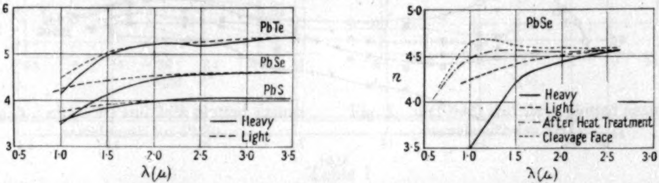


Fig. 1. Effect of surface treatments on refractive index values.

Following the work of Hogarth and Granville (1951) an attempt was made to recrystallize the amorphous layers formed when the surface was polished. The specimens were baked for 4 hours at 250° c at a pressure below 5×10^{-5} mm Hg. The effect of this treatment on the value of *n* obtained for a PbSe specimen is shown in fig. 1 (*b*). Also included are some later values obtained from a cleavage face. Comparison shows that some success was achieved. At the same time Dr. O. S. Heavens of Reading University carried out a series of electron diffraction examinations of these specimens before and after baking and showed that, though partial recrystallization occurred, it was by no means complete. Furthermore even with treatments at this relatively low temperature, appreciable amounts of material, either PbSe or free Se, were sublimed from the crystal surface on to the walls of the oven.

Comparable variations in the value of *k* were observed after these various treatments. It was interesting to note that at longer wavelengths, where for all the materials *K* has decreased to values of the order of $10^5\,\text{cm}^{-1}$ or less, and the penetration distance of the light ($1/K$ cm) has increased to appreciable values (1000 Å), the differences in *n* and *k* values caused by various treatments became quite small. It seems reasonable to suppose that provided $1/K$ cm is appreciably greater than the depth to which the material is disturbed during the surface preparation, the errors introduced from this cause will be much reduced.

However, in view of the uncertainty introduced by these processes, and since it was desired to measure n and k in regions of high absorption, the use of such surfaces was abandoned, and measurements restricted to cleavage faces wherever possible. This involved accepting certain errors which are introduced if the cleavage surfaces are not perfectly flat.

(ii) *Measurements on Cleavage Faces*

PbS. In all, some fifteen specimens of lead sulphide have been examined. The results for six different cleavage faces, and for comparison, one polished surface ('Heavy' treatment) are presented here. The variation of n with λ is shown in fig. 2, and in figs. 3(*a*) and (*b*) the real and imaginary parts of the dielectric constant are plotted against λ. The specimens came from various sources,* and included synthetic crystals grown by Mr. W. D. Lawson of this laboratory. Their electrical properties have been measured by Dr. E. H. Putley, and in table 1 below are given the sources and the concentration of impurity centres as deduced from these measurements of Hall coefficients (the centres being assumed to be fully ionized).

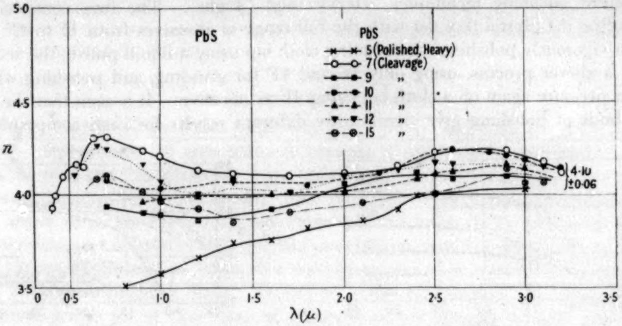


Fig. 2. Refractive index plotted against wavelength for various specimens of PbS.

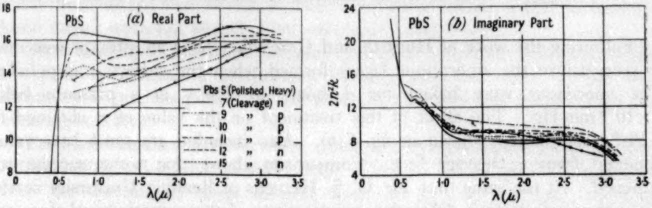


Fig. 3. $n^2(1-k^2)$ and $2n^2k$ plotted against wavelength for various specimens of PbS.

---

* Acknowledgment is made to the donors of the material.

PbSe. The real and imaginary parts of the dielectric constant for a typical cleavage face of a synthetic crystal of PbSe are shown in fig. 4.

PbTe. A similar curve for a typical specimen of PbTe is given in fig. 5.

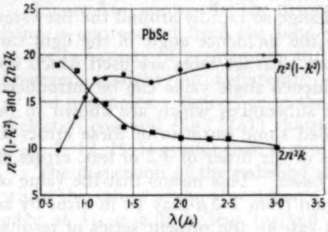


Fig. 4. $n^2(1-k^2)$ and $2n^2k$ plotted against wavelength for PbSe.

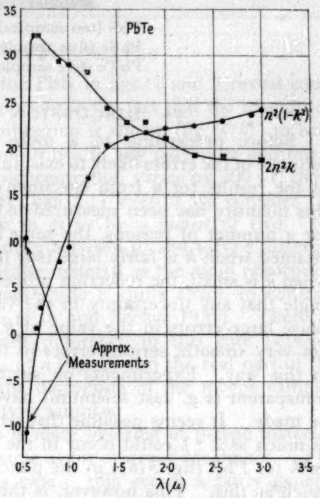


Fig. 5. $n^2(1-k^2)$ and $2n^2k$ plotted against wavelength for PbTe.

Table 1

Specimen No.	Source	Impurity concentration centres/cm³	Type
5	Sardinian Galena	6×10^{19}	p
7	Broken Hill N.S.W. (Mr. R. B. Lawrance)	$4{\cdot}5\times10^{17}$	n
9	Synthetic (Mr. W. D. Lawson)	2×10^{19}	p
10	Aberdeen (Dr. W. Paul)	$1{\cdot}3\times10^{18}$	p
11	Synthetic (Mr. W. D. Lawson)	$1{\cdot}5\times10^{18}$	p
12	Cumberland	7×10^{18}	n
15	Wisconsin U.S.A. (British Museum)	6×10^{16}	n

One feature of the investigation has been that for all three materials the values of the refractive index obtained from various samples whether polished or cleavage agreed well one with another at wavelengths of the order of 3 μ, where the absorptions have fallen to a relatively low value. This enables some reasonably precise figures for *n* at this wavelength to be given (table 2). It is possible that some slight change occurs in the refractive index of PbSe and PbTe at longer wavelengths, though measurements using a PbTe cell out to

$4{\cdot}6\,\mu$ indicate that the change is not greater than 5% (the estimated error) up to this wavelength.

Table 2

Material	n at $3\,\mu$
PbS (ten samples)	$4{\cdot}10 \pm 0{\cdot}06$
PbSe (five samples)	$4{\cdot}59 \pm 0{\cdot}06$
PbTe (five samples)	$5{\cdot}35 \pm 0{\cdot}10$

§4. DISCUSSION OF RESULTS—GENERAL

Before proceeding to a more detailed discussion of the results, some estimate of the errors likely to exist in n and k must be made. The reproducibility of the results for n from specimen to specimen of lead sulphide indicate that this quantity has been measured to an accuracy of better than 2%. However, for a number of reasons, the same order of accuracy in the value of k is only attained when k is fairly large (say $0{\cdot}4$). With a high refractive index material, when k is small, the reflection properties change so rapidly around the Brewster angle that any uncertainty in the value of the incidence angle of the light can cause large errors in the value of k. When cleavage faces are used which are not very smooth, serious errors in the incidence angle value can be introduced in this way. Experiments on surfaces of substances which are known to be transparent (e.g. cast selenium) have enabled some estimate of these errors to be made. It seems possible that when k is of the order of $0{\cdot}2$ or less, errors of as much as 2 : 1 could occur in the worst cases. This means that the value of $2n^2k$ for PbS (fig. 3 (*b*)) in the plateau region from $1\text{–}3\,\mu$ may be in error by as much as this. This however, is the worst case in the present series of results, and the errors in the value of k decrease very rapidly as k increases, the minimum in the reflection curve at the Brewster angle becoming less sharp.

It is also possible that systematic errors may have been introduced throughout the work due to oxidation of the surfaces. No check has been made on this, beyond observing that the values obtained for a single specimen do not change appreciably over several days, though it is likely that complete oxidation occurs in the inevitable ten or fifteen minutes time lag between cleaving the sample and taking the first readings.

Comparison of the values reported here for n with other data is difficult in view of the scarcity of previous reliable work. However, Drude (1890) found values of n of about $4{\cdot}0$ for polished galena at $0{\cdot}589\,\mu$. If the value of the reflection coefficient for PbS is calculated from the data, the value of $0{\cdot}39$ found from $1\,\mu$ to $3\,\mu$ agrees well with the recent measurement of $0{\cdot}40$ of Paul, Jones and Jones (1951) though the latter figure is only quoted to ± 10%.

Comparison of the absorption data with Gibson's work (1950, 1952) shows that there is agreement in the positions of the main absorption bands for each material. The value of $3{\cdot}0\text{–}3{\cdot}2\,\mu$ found here for the long wavelength absorption edge in PbS agrees with the value of $3{\cdot}0\,\mu$ quoted by Paul, Jones and Jones (1951). It is slightly greater than the $2{\cdot}7\text{–}2{\cdot}8\,\mu$ reported by Gibson (1952) but the resolution of his monochromator was only $0{\cdot}4\,\mu$. In regions where the absorption coefficient K is large (say greater than $2 \times 10^5\,\mathrm{cm}^{-1}$), there is good agreement between this work, and Gibson's values for layers. In particular for PbTe at $0{\cdot}5\,\mu$, a value of $8 \times 10^5\,\mathrm{cm}^{-1}$ is found here, as against Gibson's $7{\cdot}5 \times 10^5\,\mathrm{cm}^{-1}$.

In the regions of low absorption, comparison is really only worth while with the values of *K* found for single crystals. In the case of PbTe in the 2–3 μ region, a large difference of nearly 5 : 1 exists between this work and Gibson's, though little accuracy is claimed for the latter work.

§ 5. DISCUSSION OF RESULTS—PbS

Consideration of the results presented for PbS in figs. 2 and 3 reveal some interesting features. Perhaps the most important is the fact that the magnitude of the absorption in the long wavelength tail region is shown (fig. 3 (*b*)) to bear no direct relation to the concentration of impurities in any particular specimen. Reference to table 1 shows that whilst this concentration varies by 200 : 1, there is no variation in the value of $2n^2k$ greater than 2 : 1 even if we consider the absolute magnitude of this quantity to be in error by a factor of two. No systematic difference is revealed in this quantity between p- and n-type material.

Figure 3 (*a*) shows that the dispersion of the material as the main absorption band is approached from the long wavelength side is quite small, in contrast to the behaviour of germanium and silicon (Fan and Becker 1951). There is, however, a slight but apparently significant difference in this case between p- and n-samples. Reference to the curves of *n* (fig. 2) and the real part of the dielectric constant (fig. 3 (*a*)) shows that the dispersion of the n-type samples (Nos. 7, 12, 15) is greater than that of the p-type samples.

The dispersion of the polished samples is more like that of p-type material, whilst the values of the absorption are the same as for cleavage specimens. The edge at 3 μ is a little less marked in this case ; this seems a fairly general characteristic of polished surfaces.

The dispersion and absorption of PbS can be roughly fitted into classical dispersion theory (Wood 1933). If we represent the absorption characteristics by two bands, with peaks at 0·3 μ and 2·0 μ, and with the appropriate halfwidths, then the dispersion can be calculated to give a reasonably good approximation to the experiment if the bands have oscillator strengths of 4·0 and 0·02 respectively. This, whilst doubtless a somewhat crude approximation, gives a rough guide to the relative strengths of the two absorption processes giving rise to the main band and the long wavelength 'tail'.

§ 6. CONCLUSION

In addition to providing values for the refractive indices of PbS, PbSe and PbTe at wavelengths beyond their main absorption edges, these experiments have provided additional information about the absorption spectra of the materials. It has been shown that, for PbS at least, the magnitude of the absorption in the long wavelength 'tail' associated with photoconductivity, is apparently independent of the concentration of impurity centres in a particular specimen. Also the positions of the principal absorption edges in all three materials have been found to agree with those observed earlier in thin films by Gibson, though for PbSe and particularly PbTe, the long wavelength fall in the absorption is scarcely sharp enough to enable an 'edge' to be defined.

All three substances show relatively little dispersion on the long wavelength side of the main absorption band. The calculations carried out for PbS show that this is due in part to the existence of the 'tail' absorption, extending in the case of PbS from 1 to 3 μ, and in part to the large width of the main absorption band. It appears that samples of PbS which are n-type show slightly more dispersion than p-type specimens.

ACKNOWLEDGMENTS

My thanks are due to Dr. E. H. Putley for the Hall constant measurements, to Dr. O. S. Heavens, Reading University, for the electron diffraction work and to Dr. T. S. Moss for much helpful advice and criticism. Acknowledgment is also made to the Chief Scientist, Ministry of Supply, and to the Controller, Her Majesty's Stationery Office for permission to publish this paper.

REFERENCES

AVERY, D. G., 1951, *Proc. Phys. Soc.* B, **64**, 1087; 1952, *Ibid.*, **65**, 425.
DRUDE, P., 1890, *Wied. Ann.*, **34**, 480.
FAN, H. Y., and BECKER, H., 1951, *Semi-Conducting Materials*, ed. H. K. Henisch (London : Butterworths Scientific Publications).
GIBSON, A. F., 1950, *Proc. Phys. Soc.* B, **63**, 756; 1952, *Ibid.*, **65**, 378.
HOGARTH, C. A., and GRANVILLE, J. W., 1951, *Proc. Phys. Soc.* B, **54**, 992.
PAUL, W., JONES, D. A., and JONES, R. V., 1951, *Proc. Phys. Soc.* B, **64**, 528.
ROBERTS, V., 1952, *J. Sci. Instrum.*, **29**, 134.
WOOD, R. W., 1933, *Physical Optics* (London : Macmillan, 3rd edn.).

Inter-Relation between Optical Constants for Lead Telluride and Silicon

By T. S. MOSS

Physics Department, Telecommunications Research Establishment,
Ministry of Supply, Great Malvern

MS. received 7th August 1952

ABSTRACT. There exists a general principle that in any oscillatory or frequency responsive system the real and imaginary components of the impedance are inter-dependent. Applied to the optical properties of materials this principle means that the real and imaginary parts of the dielectric constant are inter-dependent. Sufficient information is now available on lead telluride (in a companion paper by D. G. Avery) and silicon for this principle to be used to calculate the refractive indices from the absorption data. The values obtained show good agreement with the experimental values. Estimates of the effect of temperature show that for PbTe the refractive index should *increase* on cooling, while for silicon it should *decrease*.

---

§ 1. INTRODUCTION

THE parameters generally used to express the optical properties of materials are the refractive index n and the extinction coefficient nk, where the complex refractive index $N = n(1-ik)$. The absorption constant $K = 4\pi nk/\lambda$.

From the theoretical point of view the most useful parameters are the real and imaginary parts of the dielectric constant, namely $\epsilon' = n^2(1-k^2)$ and $\epsilon'' = 2n^2k$.

Now these two parameters are not mutually independent but if, for example, one parameter is specified for all frequencies, then the other is *completely* determined. The real part of the dielectric constant at a frequency ν_1 (i.e. ϵ_1') is given in terms of the imaginary part by the integral (Bode 1945)

$$\epsilon_\infty' - \epsilon_1' = \frac{2}{\pi}\int_0^\infty \frac{\nu\epsilon'' - \nu_1\epsilon_1''}{\nu^2 - \nu_1^2}\,d\nu,$$

where ϵ_∞' is the value of the real part at infinite frequency. Now $\epsilon_\infty' = 1$, so that

$$\epsilon_1' - 1 = -\frac{2}{\pi}\int_0^\infty \frac{\nu\epsilon'' - \nu_1\epsilon_1''}{\nu^2 - \nu_1^2}\,d\nu.$$

Hence at long wavelengths where $\nu_1 \to 0$ the real part of the dielectric constant is given by

$$\epsilon_0' - 1 = -\frac{2}{\pi}\int_0^\infty \epsilon''\,\frac{d\nu}{\nu} \quad \text{or} \quad n_0^2 - 1 = \frac{2}{\pi}\int_0^\infty 2n^2k\,\frac{d\lambda}{\lambda}$$

since k is negligible compared with n for a non-metal as λ tends to ∞. This result has been given by Birnbaum (1949).

Thus from the knowledge of the complete absorption behaviour it is relatively simple to calculate the long wavelength value of the refractive index. Until recently no materials had been measured in sufficient detail for this theory to be applied, but experimental data on both silicon and lead telluride are now adequate to check the theory.

§ 2. APPLICATION TO PbTe

The absorption curve for PbTe, as determined by Avery (1951, 1953) is shown in fig. 1. The data are plotted as $2n^2k/\lambda$, so that the area under the curve gives the required integral. The long wavelength part of the curve has been plotted from results by Gibson (1952). As the absolute accuracy of the latter results is admittedly poor, they have been scaled up to match Avery's values in the 2·5–3·5 μ region.

The area under the curve is $\frac{1}{2}\pi(n_0^2-1)=46\cdot4$, giving $n_0=5\cdot5$. This figure is very near to Avery's value, namely 5·35 at 3·0 μ. Measurements at longer wavelengths show that the index is still increasing slowly with wavelength (rather than decreasing, as for a normal dispersion curve), the value being approximately 5·4 at 4·0 μ, so that the discrepancy between the observed and calculated values of n_0 must be very small.

It is of interest to try to evaluate the temperature dependence of the refractive index (as this has not so far been measured) from the temperature variation of the absorption curves. According to Gibson (1952) the long wavelength edge moves to longer wavelengths on cooling, the shift (in terms of energy) being $dE/dT=+4\cdot2\times10^{-4}$ ev/deg c. There is also a shift of the short wavelength lattice edge, estimated as $dE/dT=1\cdot5\times10^{-4}$ ev/deg. c (Gibson 1950). Hence, on cooling to liquid nitrogen temperature (77° K), the energy changes should be 0·09 ev and 0·032 ev. On the assumption that the lower value applies from the absorption peak to the beginning of the long wavelength edge (∼3·5 μ), and that the actual magnitude of the absorption is unchanged, the broken curve of fig. 1 is obtained.

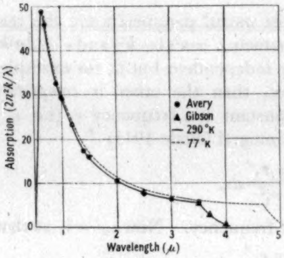


Fig. 1. Absorption in PbTe.

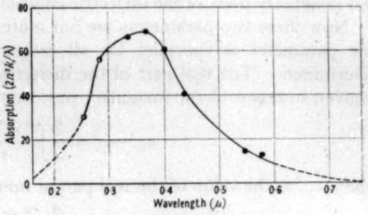


Fig. 2. Absorption in silicon.

From the area under this curve the estimated index at 77° K is given by $n_0^2-1=34\cdot6$ or $n_0=5\cdot9$. In view of the assumptions used, this value cannot be very accurate, but it is sufficient to show that a large increase in the index can be expected on cooling.

It is interesting to compare this calculated change of refractive index with the change in the wavelength of the absorption edge (which coincides with the 'threshold' wavelength of photosensitivity for PbTe) on the basis of the approximate relation found to hold for many photoconductors, namely $n^4/\lambda=$ constant (Moss 1952). From the figures obtained above $(n_{77})^4/(n_{290})^4=1\cdot33$ and $\lambda_{77}/\lambda_{290}=1\cdot35$. The good agreement obtained (in view of the assumptions made) lends support to the relation quoted above between refractive index and threshold wavelength.

§ 3. APPLICATION TO SILICON

Fairly comprehensive measurements of the optical constants of silicon have been published by Pfestorf (1926), although the results are not so complete as for PbTe. The estimated absorption curve based on these results is given in fig. 2. The long wavelength part of the curve ($\lambda > 0{\cdot}55\,\mu$) has been deduced by interpolation of n and k between Pfestorf's values and those obtained by direct transmission measurements (Moss 1952). At short wavelengths the curve has been extrapolated to have the shape of a theoretical absorption band for the imaginary part of the dielectric constant.

From the area under the curve we find $n_0{}^2 - 1 \simeq 10{\cdot}4$ or $n_0 = 3{\cdot}4$. From Briggs' (1950) measurements the experimental value (extrapolated to long wavelengths) is $n_0 = 3{\cdot}43$. Thus, again, good agreement is obtained.

A direct estimate of the change of refractive index with temperature cannot be made for silicon as the data are insufficient. It is, however, interesting to note that according to Becker and Fan (1949) the absorption edge moves to shorter wavelengths on cooling (in contradistinction to PbTe), the energy shift being $dE/dT = -4{\cdot}5$ ev/deg. c.

Consequently the refractive index would be expected to decrease on cooling, as indeed is said to be the case by Briggs (1950). An estimate of the change in refractive index to be expected can be obtained if the absorption curve is taken to resemble that of a classical oscillator, i.e. $2n^2k = A\gamma\nu/\{(\nu^2 - \nu_0{}^2)^2 + \gamma^2\nu^2\}$ where A is given by the theory as $N\boldsymbol{e}^2/\pi\boldsymbol{m}$, N being the density of oscillators, ν_0 the frequency of peak absorption, and γ the damping term.

Hence
$$n_0{}^2 - 1 = \frac{2}{\pi}\int_0^\infty \frac{A\gamma\, d\nu}{\{(\nu^2 - \nu_0{}^2)^2 + \gamma^2\nu^2\}} = A/\nu_0{}^2.$$

In terms of the quantum energy $E_0 = \boldsymbol{h}\nu_0$ we may write $n_0{}^2 - 1 = C/E_0{}^2$ where C is some constant. From fig. 2 $E_0 = 3{\cdot}4$ ev, so that $C = 12{\cdot}5$ gives the correct value of n_0 at room temperature, namely 3·43. On cooling to liquid nitrogen temperature the shift of the absorption *edge* would be 0·1 ev. If we assume that the absorption *peak* suffers the same energy shift, then E_0 becomes 3·5 ev and the refractive index at $77°\,\text{K} = 3{\cdot}34$*. Thus $(n_{77}/n_{290})^4 = 0{\cdot}90$, while for comparison the shift of the threshold wavelength is $\lambda_{77}/\lambda_{290} = 1{\cdot}14/(1{\cdot}14 + 0{\cdot}1) = 0{\cdot}92$, so that again a correlation between the threshold wavelength and the fourth power of the refractive index is indicated.

§ 4. CONCLUSIONS

It is shown that for lead telluride and for silicon the values of the long wavelength index of refraction calculated from the absorption data agree well with the measured values; thus there can be no marked absorption bands further into the ultra-violet.

From the temperature dependence of the absorption the temperature variation of refractive index is estimated. It is shown that the index should increase for PbTe and decrease for silicon on cooling, and the rough estimate of the magnitude of the change is found to be compatible with the empirical relation

---

* The change of 0·2% in N due to lattice contraction is ignored.

(threshold wavelength) ∝ (refractive index)^4. However, in view of the assumptions made, the close agreement with this relation must be to some extent fortuitous.

ACKNOWLEDGMENTS

Acknowledgment is due to the Chief Scientist, Ministry of Supply, and the Controller, Her Majesty's Stationery Office, for permission to publish this paper, and to Dr. Avery for supplying the data on lead telluride.

REFERENCES

AVERY, D. G., 1951, *Proc. Phys. Soc.* B, **64,** 1087; 1953, *Ibid.*, **66,** 134.
BECKER, M., and FAN, H. Y., 1949, *Phys. Rev.*, **76,** 1531.
BIRNBAUM, G., 1949, *Phys. Rev.*, **76,** 178.
BODE, H. W., 1945, *Network Analysis and Feedback Amplifier Design* (New York: Van Nostrand).
BRIGGS, H. B., 1950, *Phys. Rev.*, **77,** 287.
GEBBIE, H. A., 1952, *Thesis*, University of Reading.
GIBSON, A. F., 1950, *Proc. Phys. Soc.* B, **63,** 756; 1952, *Ibid.*, **65,** 378.
MOSS, T. S., 1952, *Photoconductivity in the Elements* (London: Butterworths Scientific Publications).
PFESTORF, G., 1926, *Ann. Phys., Lpz.*, **81,** 906.

APPENDIX

Application to Germanium

Recent measurements of $2n^2k$ over the wavelength range 0·25–1·0 μ by Gebbie (1952) have been integrated as indicated above, the calculated value of the refractive index being $n_0 \simeq 3{\cdot}8$. In view of the somewhat tentative nature of the experimental data, the agreement with the accepted value of $n_0 = 4{\cdot}0$ is reasonable.

Diffraction Images in Systems with an Annular Aperture

By E. H. LINFOOT and E. WOLF*
The Observatories, University of Cambridge

MS. received 23rd September 1952

ABSTRACT. The paper is concerned with the effects of central obstruction of the aperture on the three-dimensional light distribution near the focus of an aberration-free optical system. Diagrams are given of the isophotes (lines of equal light intensity) in two selected special cases and are compared with the corresponding diagram for an unobstructed aperture. It appears that when the central obstruction is large the bright central nucleus of the three-dimensional image becomes longer and narrower, so that focal depth and resolving power are both increased. A central obstruction of ratio 0·25, on the other hand, is found to have practically no effect on the size and shape of the bright nucleus.

§1. INTRODUCTION

The intensity distribution in space near the geometrical focus of an error-free pencil of monochromatic light bounded by a circular aperture has been given in a diagram by Zernike and Nijboer (1949). The diagram exhibits in a striking way a peculiarity of the diffraction image already noted by Dennis Taylor (1893), namely the quasi-tubular structure of the luminous core of the image. It is this tubular elongation of the bright central nucleus of the diffraction image, exceeding what we could expect on the basis of more elementary considerations, which explains the excellent performance of 6-in. or 8-in. refracting telescopes in spite of their considerable secondary spectrum.

In reflecting telescopes there is no secondary spectrum. Colour effects arise from the fact that the diffraction patterns surrounding the geometrical focus are of different sizes in different wavelengths, but these are so inconspicuous that their existence is usually ignored even in the literature of diffraction theory †. However, in most reflecting telescopes there is a different complication which affects the diffraction images, namely the central obstruction (usually circular in form) of the aperture by the diagonal flat or by the secondary mirror.

It seems unlikely that anyone who has interested himself in this subject should not feel a strong desire to know what is the effect of a central obstruction of the aperture on the three-dimensional light distribution near focus. In the present note this question is discussed by a straightforward application of Lommel's classical formulae (Lommel 1885)‡, and diagrams of the light distribution are given, in figs. 4 and 5 respectively, for two selected values of the linear obstruction ratio ϵ, namely, $\epsilon = 0{\cdot}25$ and $\epsilon = 0{\cdot}707$. The corresponding diagram for the unobstructed aperture is given for comparison in fig. 3; it is substantially identical with Zernike and Nijboer's, which was obtained by a different method. An earlier diagram by Berek (1926), based, like ours, on Lommel's results, contained inaccuracies.

---

* Now at the Department of Applied Mathematics, University of Edinburgh.

† They are considered by Mecke (1920) and Picht (1931). A more detailed discussion, on somewhat different lines, is contained in our paper (Linfoot and Wolf 1953) on telescopic star images.

‡ Some extensions of Lommel's tables of U_1 and U_2 were needed for figs. 4 and 5, and these were kindly prepared for us by the Cambridge University Mathematical Laboratory. Since this paper was written, A. Boivin, 1952, *J. Opt. Soc. Amer.*, **42,** 60, has published interesting new expansions for the Lommel functions with arguments $u'=\epsilon^2 u$, $v'=\epsilon v$ (see eqn. (4)) and applied them to the discussion of diffraction by arrays of ring-shaped apertures.

§ 2. ANALYTICAL FORMULAE

We use substantially the same notation and approximations as in Linfoot and Wolf (1953). For convergent spherical waves of unit amplitude, issuing from a circular aperture of radius R and having radius of curvature $f = \mathrm{CO}$ (fig. 1)

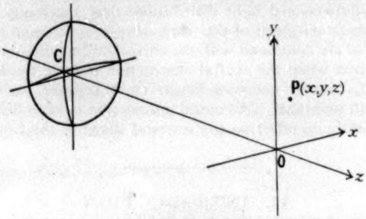


Fig. 1.

at the moment of emergence, the complex displacement at a point P(x, y, z) in space near the geometrical focus 0 is given by the equation

$$u_\lambda^{(R)}(\mathrm{P}) = \frac{ikR^2}{f}[\exp\{ik(f-\mathrm{CP})\}]\int_0^1 [\exp(\tfrac{1}{2}iu\rho^2)]\,J_0(v\rho)\,\rho\,d\rho, \quad \ldots\ldots(1)$$

where

$$u = kR^2z/f^2, \quad v = kRr/f, \quad k = 2\pi/\lambda, \quad r = +(x^2+y^2)^{1/2}. \quad \ldots\ldots(2)$$

To allow for the effect of a central obstruction of radius $R' = \epsilon R$, we subtract from $u_\lambda^{(R)}(\mathrm{P})$ the complex quantity

$$u_\lambda^{(R')}(\mathrm{P}) = \frac{ikR'^2}{f}[\exp\{ik(f-\mathrm{CP})\}]\int_0^1 [\exp(\tfrac{1}{2}iu'\rho^2)]\,J_0(v'\rho)\,\rho\,d\rho, \quad \ldots\ldots(3)$$

in which

$$u' = kR'^2z/f^2 = \epsilon^2 u, \quad v' = kR'r/f = \epsilon v. \quad \ldots\ldots(4)$$

The intensity at P is then the squared modulus of the quantity

$$\begin{aligned} u_\lambda(\mathrm{P}) &= u_\lambda^{(R)}(\mathrm{P}) - u_\lambda^{(R')}(\mathrm{P}) \\ &= \frac{ikR^2}{f}[\exp\{ik(f-\mathrm{CP})\}]\Big[\int_0^1 [\exp(\tfrac{1}{2}iu\rho^2)]\,J_0(v\rho)\,\rho\,d\rho \\ &\qquad - \epsilon^2\int_0^1 [\exp(\tfrac{1}{2}iu'\rho^2)]\,J_0(v'\rho)\,\rho\,d\rho\Big]. \quad \ldots\ldots(5) \end{aligned}$$

Now

$$\int_0^1 [\exp(\tfrac{1}{2}iu\rho^2)]\,J_0(v\rho)\,\rho\,d\rho = \frac{1}{u}[\exp(\tfrac{1}{2}iu)][U_1(u,v) - iU_2(u,v)]$$

$$\int_0^1 [\exp(\tfrac{1}{2}iu'\rho^2)]\,J_0(v'\rho)\,\rho\,d\rho = \frac{1}{u'}[\exp(\tfrac{1}{2}iu')][U_1(u',v') - iU_2(u',v')],$$

where $U_1(u,v)$, $U_2(u,v)$ are Lommel functions. Thus (5) can be written

$$\begin{aligned} u_\lambda(\mathrm{P}) = \frac{ikR^2}{f}[\exp\{ik(f-\mathrm{CP})\}]\Big[&\frac{1}{u}[\exp(\tfrac{1}{2}iu)](U_1 - iU_2) - \frac{1}{u}[\exp(\tfrac{1}{2}i\epsilon^2 u)] \\ &\times (U_1' - iU_2')\Big], \quad \ldots\ldots(6) \end{aligned}$$

where, U_1, U_2, U_1', U_2' are written for $U_1(u,v)$, $U_2(u,v)$, $U_1(u',v')$, $U_2(u',v')$ respectively, and the intensity $I_\lambda(\mathrm{P})=|u_\lambda(\mathrm{P})|^2$ is given by the equation

$$\begin{aligned} I_\lambda(\mathrm{P}) &= \frac{k^2R^4}{f^2u^2}\left|[\exp(\tfrac{1}{2}iu)](U_1-iU_2)-[\exp(\tfrac{1}{2}i\epsilon^2u)](U_1'-iU_2')\right|^2 \\ &= \frac{\pi^2R^4}{\lambda^2f^2}\{M^2(u,v)-2\epsilon^2N(u,v;\,u',v')+\epsilon^4M^2(u',v')\}, \end{aligned} \qquad \ldots\ldots(7)$$

where

$$\left.\begin{aligned} M^2(u,v) &= \left(\frac{2}{u}\right)^2(U_1^2+U_2^2),\quad M^2(u',v')=\left(\frac{2}{u'}\right)^2(U_1'^2+U_2'^2) \\ N(u,v;\,u',v') &= \frac{4}{uu'}[(U_1U_1'+U_2U_2')\cos\tfrac{1}{2}(1-\epsilon^2)u \\ &\qquad +(U_2U_1'-U_1U_2')\sin\tfrac{1}{2}(1-\epsilon^2)u]. \end{aligned}\right\} \qquad \ldots\ldots(8)$$

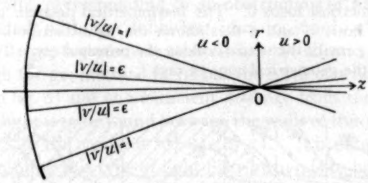


Fig. 2.

On the axis of the converging pencil, $v=0$ and (6) gives for the intensity the expression

$$I_\lambda(z,0)=\frac{4\pi^2R^4}{\lambda^2f^2}\left(\frac{\sin\frac{1}{4}u(1-\epsilon^2)}{\frac{1}{4}u}\right)^2. \qquad \ldots\ldots(9)$$

Thus the central obstruction increases the distance between the zeros along the axis of the system by a factor $1/(1-\epsilon^2)$ but they remain equally spaced—at least in the range where our approximations are valid, namely the part of the axis where $u/4\pi$ does not become large compared with unity.

In the geometrical focal plane, $u=0$ and the expression (7) for the intensity reduces to

$$I_\lambda(0,r)=\frac{4\pi^2R^4}{\lambda^2f^2}\left[\frac{2J_1(v)}{v}-\epsilon^2\frac{2J_1(\epsilon v)}{\epsilon v}\right]^2. \qquad \ldots\ldots(10)$$

The zeros of this function give, in v-units, the radii of the 'Airy dark rings' corresponding to a centrally obstructed aperture. The expressions (9) and (10) were given (with a different normalization) many years ago by Steward (1925).

§ 3. COMPUTATIONAL RESULTS

To interpret figs. 3, 4 and 5 we note that each of them represents one quarter of a bisymmetrical pattern obtained by reflecting it in the u- and v-axes. This bisymmetrical pattern shows the isophotes (lines of equal light intensity) near focus in any meridional section of the pencil. The diagrams apply to pencils of all sufficiently long focal ratios (to those, in fact, for which Lommel's formulae are valid approximations in the sense of our paper (Linfoot and Wolf 1953)) and the use of $(u,\ v)$ coordinates in the diagrams is equivalent to scale-normalizing the cylindrical cartesian coordinates $(r,\ z)$ in accordance with equations (2). The intensity at the geometrical focal point (0, 0) is normalized to unity in each figure.

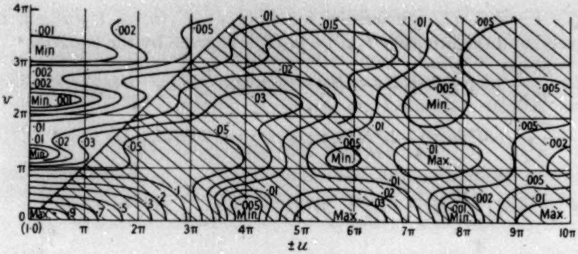


Fig. 3. Isophotes near focus of an aberration-free pencil without central obstruction. The intensity at the focus (0, 0) is normalized to unity. The scale-normalized coordinates (u, v) possess physical interpretations; $u/4\pi$ is the number of fringes of defocusing, v/π the number of fringes of lateral displacement of the point (z, r) from the geometrical focus 0. The bisymmetrical diagram obtained by reflecting the figure in both u- and v-axes shows the light-distribution in any meridional section of the pencil; the u-axis is along the principal ray. The shaded area shows the region of the geometrical cone of rays

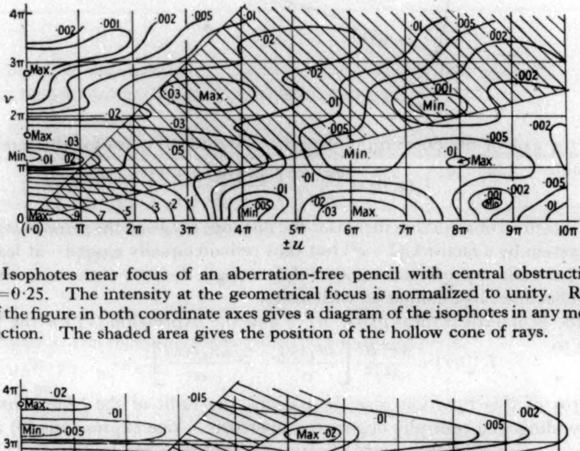


Fig. 4. Isophotes near focus of an aberration-free pencil with central obstruction ratio $\epsilon = 0{\cdot}25$. The intensity at the geometrical focus is normalized to unity. Reflection of the figure in both coordinate axes gives a diagram of the isophotes in any meridional section. The shaded area gives the position of the hollow cone of rays.

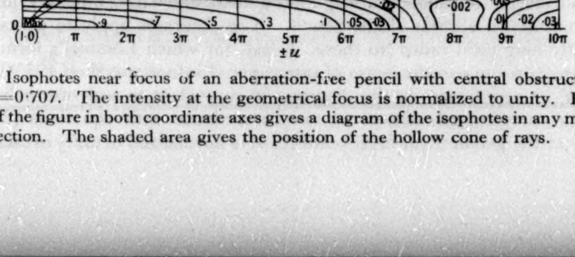


Fig. 5. Isophotes near focus of an aberration-free pencil with central obstruction ratio $\epsilon = 0{\cdot}707$. The intensity at the geometrical focus is normalized to unity. Reflection of the figure in both coordinate axes gives a diagram of the isophotes in any meridional section. The shaded area gives the position of the hollow cone of rays.

The shaded areas in the bisymmetrical patterns show where the geometrical light-cones meet the meridional (u, v)-plane in each case. In the unobstructed case (fig. 3) the cone of rays is a solid cone, with axis lying along the u-axis and the lines $|v/u| = 1$ lie in its surface. In the centrally obstructed cases the cone is hollow; its axis lies along the u-axis and its 'body' lies between the two conical surfaces traced out when the lines $|v/u| = \epsilon$ and $|v/u| = 1$ are rotated about the u-axis (see fig 2). In fig. 4, $\epsilon = 0{\cdot}25$, in fig. 5, $\epsilon = 0{\cdot}707$.

The value $\epsilon = 0{\cdot}25$ was selected for computation because it corresponds to the greatest central obstruction which is regarded as tolerable by users of visual reflecting telescopes. A comparison of figs. 3 and 4 shows how small is the effect of this obstruction on the relative intensities in different parts of the image. In particular, the size and shape of the bright central nucleus is almost unaffected. There is, however, an increase in the intensity of the first Airy bright ring, the effects of which might occasionally be visible to a keen observer.

The value $\epsilon = 0{\cdot}707$, corresponding to an obstruction of half the area of the aperture, was selected with the following idea in mind. When the central obstruction is fairly large, for example when it has the above value, the pencil of rays passing through the geometrical focus has the form of a hollow circular cone (shown in section in fig. 5) and at a sufficient distance from the geometrical focal plane almost all the light is to be found between the walls of this cone, that is to say, in the notation of § 2, in the region $0{\cdot}707 < |v/u| < 1$. It is natural to ask whether anything at all resembling this type of light distribution is to be found in the near neighbourhood of the geometrical focus. Figure 5 provides an answer to the question: close to focus, the light distribution has the same general character as in the case of the unobstructed aperture and is strikingly different from that predicted by geometrical optics. The main effect, in this region, of the obstruction is to draw out the central nucleus of the image along the axis of the system to approximately twice its former length, its cross section being correspondingly reduced.

These results may be of some practical value in connection with the design of lens–mirror systems, in which the use of mirrors generally involves central obstruction of the beam, while the refractive elements usually introduce some chromatic variation of focus. The existence and dimensions of a tubular core to the diffraction image have here, just as in the case of telescope doublets, an important bearing on the amount of chromatic variation of focus which can be tolerated in the system (Conrady 1923, Linfoot and Wolf 1953).

More academic, but perhaps not entirely without interest, is the point that a large central stop on the objective of a refracting telescope not only increases resolving power by decreasing the lateral diameter of the bright central nucleus of the image but also, by elongating the nucleus in the axial direction, reduces the disturbing effects of chromatism on its colour-composition at best focus.

REFERENCES

Berek, M., 1926, *Z. Phys.*, **40,** 421.

Conrady, A. E., 1923, *Dictionary of Applied Physics*, Vol. IV (London: Macmillan), p. 222.

Linfoot, E. H., and Wolf, E., 1953, *Mon. Not. R. Astr. Soc.*, **112,** 452.

von Lommel, E., 1885, *Abh. der Bayerischen Akad.*, **53,** 233.

Mecke, R., 1920, *Ann. Phys., Lpz.*, (4) **61,** 471; **62,** 623.

Picht, J., 1931, *Optische Abbilding* (Braunschweig: Vieweg), p. 102.

Steward, G. C., 1925, *Phil. Trans. Roy. Soc.* A, **225,** 131.

Taylor, H. Dennis, 1893, *Mem. Roy. Astr. Soc.*, **54,** 77.

Zernike, F., and Nijboer, B. R. A., 1949, Contribution to *Théorie des images optiques* (Paris: Éditions de la Revue d'Optique), p. 227.

LETTER TO THE EDITOR

On the Decay of Radio Echoes from Meteor Trails

The interpretation of the long duration echoes observed from meteor trails has been the subject of some dispute. In recent papers, Kaiser and Closs (1952) and Greenhow (1952) have given a comprehensive theory to explain the nature of the scattering process, the duration of the echo, and fluctuations in amplitude of the radio echo with time. In this theory diffusion is considered to be the main process reducing the volume density of electrons in the trail and recombination is neglected. Feinstein (1952) has now criticized this interpretation on the grounds that the observed value for the recombination coefficient in the E region of the ionosphere ($\alpha = 10^{-8}$ cm^3 sec^{-1}) should also apply to the conditions in a meteor trail. In this case the duration of echoes from trails of line density in excess of about 10^{13} electrons/cm when observed on a wavelength of 8 m would be limited to about 8 sec, whereas on the diffusion theory the duration may be much greater and proportional to the initial line density.

The discrepancy between the high value of α observed in the E region and the theoretical value for radiative recombination (10^{-12} cm^3 sec^{-1}) has been explained on the assumption of electron loss through attachment, and, more recently, through dissociative recombination which is important (Bates 1950) under certain conditions. The former explanation is now regarded as the least likely, but in any case such attachment would not be significant for the electron and molecular densities in meteor trails at a height of 90 km. Dissociative recombination might be important if molecular ionization was predominant in meteor trails. However, the fact that the ionization potentials of the meteor atoms are lower than those of the atmospheric constituents, and the absence of spectral evidence for ionized lines other than those of meteor atoms, suggests that the ionization is mainly atomic. Hence there are no strong theoretical grounds for anticipating that the value of α found in the E region should apply to meteor trails. Evidence in support is provided by the observed variation of radio echo duration with echo amplitude and with visual magnitude (Greenhow and Hawkins 1952) for line densities up to 10^{15} cm^{-1}. These results indicate that the effective recombination coefficient for meteor trails is not greater than 5×10^{-10} cm^3 sec^{-1}.

In Feinstein's hypothesis concerning long duration echoes the E region value of α is retained and initial shock wave expansion of the trail is suggested as a mechanism for dispersing the high electron density before recombination becomes effective. As a consequence he concludes that the deep amplitude fluctuations are produced by interference between waves reflected at different levels in the cylinder of ionization. It is clear that such interference fluctuations could occur only after the column has expanded sufficiently to reduce the electron density below the critical value; thus, unless the boundary is exceedingly sharp, the amplitude of the echo during this period would be very much smaller than during the initial phase. Long duration echoes observed on a wavelength of 8 m frequently exhibit fluctuation periods of the order of 0·1 second. On Feinstein's hypothesis, such an echo of 60 sec duration would require a meteor trail in which the electron density decreases from its maximum value to zero in a boundary layer of a few metres thickness and which retains this structure until its radius is of the order of 1 km. A more serious objection is that the fluctuations would not commence simultaneously when observed on different wavelengths, a prediction which is in contradiction with observation. The simple periodic fluctuations with deep cusp-like minima predicted by Feinstein are rare; more commonly the echo is complex. When they occur, a continuous gradation into a Rayleigh distribution of amplitudes is observed, the mean fluctuation period remaining constant throughout. In addition the mean amplitude is generally substantially constant over many fluctuation periods. These facts show many points of disagreement with Feinstein's predicted interference effects, but are explicable by the theories of Kaiser and Closs and of Greenhow in which the fluctuations are attributed to interference effects arising from trail distortion due to winds in the upper atmosphere.

We agree that the heat energy released by the meteor may enhance the initial rate of expansion but it would appear that discussion of this question must await further theoretical developments. In our interpretation of the experimental evidence such effects have only a small influence on the important echo characteristics, particularly the duration.

Jodrell Bank Experimental Station,
University of Manchester.
28th November 1952.

T. R. KAISER.
J. S. GREENHOW.

BATES, D. R., 1950, *Phys. Rev.*, **77**, 718, **78**, 592.
FEINSTEIN, J., 1952, *Proc. Phys. Soc.* B, **65**, 741.
GREENHOW, J. S., 1952, *Proc. Phys. Soc.* B, **65**, 169.
GREENHOW, J. S., and HAWKINS, G. S., 1952, *Nature, Lond.*, **170**, 355.
KAISER, T. R., and CLOSS, R. L., 1952, *Phil. Mag.*, **43**, 1.

BOOK NOTICES

Hydraulic Research in the United States, edited by H. K. MIDDLETON and S. W. MATCHETT. National Bureau of Standards Miscellaneous Publication 205. Pp. xii + 200. (Washington, D.C.: U.S. Department of Commerce, 1952.) $1.00.

Electrolyse, Report of a conference organized by the Centre National de la Recherche Scientifique. Pp. 147. (Paris: Centre National de la Recherche Scientifique, 1952.) 1500 fr.

Contribution à l'étude expérimentale des écoulements gazeux rapides par les méthodes de perturbation utilisation des ultrasons, by GENEVIÈVE DUBOIS. Pp. iii + 57. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air (No. 269), 1952.) 650 fr.

CONTENTS OF SECTION A

PAGE

Dr. Suraj N. Gupta. Quantum Electrodynamics with Auxiliary Fields . . . 129

Dr. L. A. Radicati. Isotopic Spin and Coulomb Forces 139

Dr. W. B. Bonnor. Certain Exact Solutions of the Equations of General Relativity with an Electrostatic Field 145

Dr. D. ter Haar and Mr. C. D. Green. The Statistical Aspect of Boltzmann's H-Theorem. 153

Dr. J. C. Barton. East-West Asymmetry of Moderate-Energy Neutrons in the Cosmic Radiation 160

Prof. L. F. Bates and Mr. D. H. Martin. Domains of Reverse Magnetization . 162

Mr. B. T. Price, Mr. D. West, Mr. J. Becker, Mr. P. Chanson, Mr. E. Nageotte and Mr. P. Treille. Further Measurements of the Ionization by Energetic Cosmic-Ray μ-Mesons 167

Dr. A. Dalgarno and Dr. H. N. Yadav. Electron Capture—II: Resonance Capture from Hydrogen Atoms by Slow Protons 173

Dr. J. S. Plaskett. The Theory of the Thomas–Fermi Electron Density . . . 178

Letters to the Editor:

- Dr. R. F. Barrow, Mr. G. Drummond and Mr. W. R. S. Garton. Ultra-Violet Bands associated with Germanium 191
- Mr. P. W. Roberts. Some Comments on the Resolution of Scintillation Spectrometers 192
- Prof. E. W. Titterton and Mr. T. A. Brinkley. Excitation Functions for the (γp) and (γT) Reactions in Lithium-7 for Energies up to 24 mev . . 194
- Mr. M. L. G. Redhead. The Production of Bremsstrahlung in Electron–Electron Collisions 196
- Dr. J. S. Kouvelites. On the Motion of Electrons in Non-sinusoidal Periodic Fields 197

Reviews of Books 199

Contents of Section B 200

Fig. 1. Surfaces of cast rods. 4 rods of tin–silver 0·1 %, 1 rod of tin.

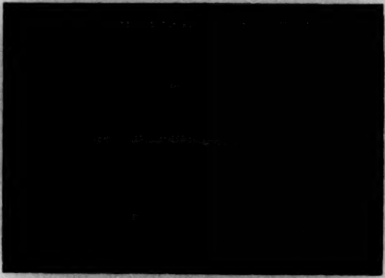


Fig. 2. Fine lines on unetched crystal surfaces (tin–silver 0·1 %. Rate 0·5 mm/min).

Fig. 3. Dendritic surfaces (tin–antimony 0·1 %. Rate 20 mm/min).

Fig. 1. Single crystal diffraction pattern from a {110} face of Hg_2Cl_2. Beam ⊥ to film. TlCl calibration pattern superimposed.

Fig. 2. Diffraction pattern from a polycrystalline portion of the film. Beam ⊥ to film. {110} planes parallel to the substrate.

Fig. 3. Two single crystals with an angle of 72° between the c_0 axis, {110} faces parallel to the substrate. Beam ⊥ to film (see figs. 8 and 9).

Fig. 4. Diffraction pattern from a twinned portion of the film. Beam ⊥ to the film.

Fig. 5. Two crystals slightly rotated on a {110} cleavage face about an axis ∥ to the beam, TlCl calibration pattern superimposed.

Fig. 6. Diffraction pattern from a crystal showing rotational slip (see fig. 9).

The Formation of Hydrogen Peroxide on Fresh Metal Surfaces

By L. GRUNBERG

Mechanical Engineering Research Laboratory, Department of Scientific and Industrial Research.

Communicated by F. T. Barwell; MS. received 31*st October* 1952

Abstract. The recent discovery by Kramer that freshly exposed metal surfaces emit electrons is considered to be the cause of the Russell effect and of Shaw's mechanical activation. Electrons originating from the surface initiate reactions involving free radicals. With water, in the presence of oxygen, this leads to the formation of hydrogen peroxide.

Metals were cut under water containing oxygen and it was found that, under similar conditions, zinc, aluminium, magnesium and nickel yield approximately the same quantity of H_2O_2. Iron gives ferric ions and copper only very little H_2O_2. During cutting, considerable quantities of hydroxides become suspended in the solution. With zinc more H_2O_2 is formed per cut in the early stages of the experiment, owing to the exposure of fresh surface by the dissolution of $Zn(OH)_2$. In the later stages each zinc atom exposed is responsible for the formation of approximately 1·5 molecules of H_2O_2. The decomposition of H_2O_2 in the presence of chips of zinc is found to be a pseudo-unimolecular process. The cutting of copper in a solution of ferrous ions yielded ferric ions. The amount of ferric ions produced seems to depend on the quantity of oxygen available in the solution.

§1. INTRODUCTION

THE reactivity of freshly formed metal surfaces is of great importance in the study of many processes, such as lubrication, corrosion and catalysis. It is normally assumed that the reactivity of such surfaces is high and resides in so-called 'active centres', but very little is known definitely about these centres. This note reports on a part of the research work on the properties of fresh metal surfaces at present in progress in the Mechanical Engineering Research Laboratory. The research work was stimulated by the discovery, made by Kramer (1949, 1950 a, b, 1951), that abraded metal surfaces emit electrons at relatively low temperatures. The 'Kramer effect' is probably responsible for two particular phenomena, which so far have found no satisfactory explanation. One of these is the 'Russell effect' discovered in 1897 (Russell 1897, 1898, 1899, 1906, 1908) and the other is the so-called 'mechanical activation' discovered by Shaw (1948). The work reported here is concerned with linking the last two phenomena, and assumes the Kramer effect to be the cause of the processes concerned. The three effects will now be briefly discussed.

1.1. *The Kramer Effect*

When freshly abraded metal specimens are exposed to what can be described as a modified Geiger–Müller counter, a definite count can be obtained. This 'activity' decreases with time and can definitely be associated with the freshly exposed metal surface. Kramer's results were confirmed by Louw and

Naudé (1949). Kramer explained the phenomenon by the existence of a particular state of metals which he described as 'non-metallic'. The mechanism of electron emission is not stated in detail and from the experiments by Haxel, Houterman and Seeger (1951) it appears that adsorption may play an important part in inducing emission. In this connection must be mentioned the work by Fianda and Lange (1951) who found that the work function of metals abraded *in vacuo* was lowered considerably by adsorption of oxygen. It thus appears that the existence of the Kramer effect can be accepted, but contrary to his explanation, the effect may be caused by heterogeneous processes on the surface of the freshly exposed metal. This problem is at present being investigated in the Mechanical Engineering Research Laboratory, using films of evaporated metals as the source of the Kramer effect. For a discussion of the Russell effect and mechanical activation the mechanism of the Kramer effect is of little importance. It seems sufficient to accept the fact that a freshly abraded metal surface is a potential source of free electrons.

1.2. *The Russell Effect*

Russell (1897, 1898, 1899, 1906, 1908) discovered that freshly abraded surfaces of certain metals produced an image on photographic plates. The 'activity' of these surfaces decreases with time, and is enhanced by temperature, as is the Kramer effect. Russell was able to show that the image produced on a photographic plate was caused by a gas and that extremely small quantities of hydrogen peroxide vapours acted in a way similar to the gas produced on freshly abraded metal surfaces. The Russell effect received a great deal of attention during the first decade of this century. A summary of the investigations has been given by Keenan (1920). The sometimes contradictory evidence seems to indicate that the presence of both oxygen and water vapour is required to produce the effect. More recently Churchill (1939) showed that hydrogen peroxide is formed on freshly abraded surfaces and that the amount formed on a cathode is very much greater than that formed on an anode. Mention must also be made of early attempts to develop a peroxide theory of corrosion (Dunstan *et al.* 1905).

Work at the Mechanical Engineering Research Laboratory confirmed most of the results by previous workers on the Russell effect. In addition, it was established that the effect can also be produced with films of evaporated metals (Grunberg and Wright 1952). A satisfactory explanation of the Russell effect can be obtained, if one assumes the Kramer effect to be the primary cause of the reaction. If free electrons are available either as direct emission or by a lowering of the work function then, in the presence of oxygen and water vapour, the following reactions are likely to occur:

$$e + H_2O \rightarrow H_2O. \rightarrow H. + (:OH)^- \qquad \ldots\ldots(1)$$

$$H. + O_2 \rightarrow HO_2. \qquad \ldots\ldots(2)$$

$$HO_2. + H_2O \rightarrow H_2O_2 + .OH \qquad \ldots\ldots(3)$$

$$e + .OH \rightarrow (:OH)^- \qquad \ldots\ldots(4)$$

Churchill (1939) suggested a scheme, which involved hydrogen ions in two of the steps. At the concentration of water vapour at which the Russell effect can be observed the number of hydrogen ions present cannot be sufficiently great to account for the intensity observed.

There is a discrepancy between the Russell and Kramer effects. Whilst the latter can be observed with relatively electronegative metals (e.g. gold) the former has not been observed with iron or copper. This can be explained by the fact that these metals decompose the hydrogen peroxide before it can react with the photographic plate. Saeland (1908) showed that interposition of copper filings or of a brass screen inhibits the Russell effect. The decomposition of hydrogen peroxide probably occurs on the surface of the metal and follows a course analogous to the ferrous ion catalysed reaction suggested by Haber and Weiss (1934, 1935):

$$H_2O_2 + e \rightarrow (:OH)^- + .OH. \qquad \ldots\ldots(5)$$

The electron involved in reaction (5) may either be a Kramer electron or may derive from the 3d band as suggested by Dowden and Reynolds (1950).

1.3. *Mechanical Activation*

During a study of the action of fluids used for lubrication in metal cutting Shaw (1948) observed that certain organic compounds react vigorously with the metal cut, although these compounds are relatively inert to the metal in bulk. Certain reactions, such as the preparation of Grignard reagents, which normally require the most anhydrous conditions, proceed vigorously even in the presence of water. Shaw termed this 'mechanical activation' and considered that the high temperatures and pressures at the tool face during cutting, together with the great reactivity of the highly stressed nascent surface produced, are responsible for the high reaction rates observed. Shaw did not discuss the properties of the 'nascent' metal surface. It seems likely that this is another manifestation of the Kramer effect, as some of the reactions studied by Shaw can be explained by electron transfer. If one assumes a partially ionic character of the alkyl- or aryl-halogen bond the formation of a Grignard compound proceeds by electron transfer according to

$$R^+I^- + e \rightarrow R^{\cdot}I^- \rightarrow R^{\cdot} + I^- \qquad \ldots\ldots(6)$$

and

$$R^{\cdots} + I^- + Mg^{2+} \rightarrow RMgI. \qquad \ldots\ldots(7)$$

Another reaction studied by Shaw was the decomposition of carbon tetrachloride during the cutting of aluminium. He observed the formation of a dark red substance which could be resolved into aluminium chloride and hexachloroethane. Considering again electron transfer to a partially ionic bond, then

$$Cl_3C^+Cl^- + e \rightarrow Cl_3C^{\cdot}Cl^- \rightarrow Cl_3C^{\cdot} + Cl^- \qquad \ldots\ldots(8)$$

and

$$Al^{3+} + 3Cl^- \rightarrow AlCl_3 \qquad \ldots\ldots(9)$$

$$2Cl_3C^{\cdot} \rightarrow C_2Cl_6 \qquad \ldots\ldots(10)$$

Shaw stated that relatively high cutting speeds (i.e. high surface temperatures) are required for the reactions to occur. The cause of this may be the relative inefficiency of the first step of reaction (8). The Kramer effect is greatly enhanced by an increase in surface temperature and then provides a sufficient number of electrons for reaction (8) to proceed.

§ 2. EXPERIMENTAL

In the previous section it was suggested that the Russell effect and Shaw's mechanical activation are caused by the same phenomenon, namely the emission of electrons from freshly exposed metal surfaces (the Kramer effect). As a preliminary to the study of the latter it seemed important to investigate whether the Russell effect—the formation of hydrogen peroxide on fresh metal surfaces—could be produced by the cutting of metals in the presence of water and oxygen.

2.1. *Apparatus and Procedure*

A metal block approximately 4½ in. × 3 in. × 2 in., is firmly attached to the bottom of a plastic trough. The trough is clamped into the machine vice of a hydraulic shaping machine and filled with 1000 ml. distilled water. The capacity of the trough is such that the metal block is completely immersed in the water. A high-speed steel tool is introduced and orthogonal cuts are taken with a feed of 0·002 in. In order to keep the temperature as low as possible, low cutting speeds of about 2 cm/sec (4 ft/min) are used except when cutting nickel, when the cutting speed was of about 2·8 cm/sec (5·5 ft/min). The cutting chips are kept submerged and occasionally rearranged in the trough to prevent accumulation and exposure to air. During the experiment the temperature of the water is measured. After taking the required number of cuts, the water is immediately removed from the trough by means of a syphoning arrangement. Aliquot portions of the water are analysed.

2.2. *Materials and Methods of Analysis*

The samples of aluminium, magnesium, nickel and zinc were of the highest purity commercially available. The sample of copper was high conductivity grade and the sample of iron was Armco iron.

The water was freshly distilled in a Pyrex apparatus and had a dissolved oxygen content of approximately 15 ml/l. All the reagents used were of A.R. quality. Hydrogen peroxide was determined in a sample of 200 ml. by iodometry in the presence of sulphuric acid. Ferric ions were determined by a similar method and also colorimetrically as thiocyanate.

2.3. *Amount of Hydrogen Peroxide formed with Different Metals*

Cutting experiments lasting one hour were carried out with zinc, aluminium, magnesium, nickel, copper and iron. 450 cuts were made, except for nickel (655 cuts). In all cases, except when cutting nickel, the water remained at laboratory temperature. With nickel the temperature of the water rose by about 11°c. The results of the experiments are shown in table 1.

Table 1. Showing the Amount of Hydrogen Peroxide present after Cutting for One Hour

Metal	No. of cuts	Dimensions of cut		Cutting speed (cm/sec)	Temp. of water (°c)	mg H_2O_2
		Length (cm)	Width (cm)			
Zinc	450	8·5	1·3	2·12	19	2·75
Aluminium	450	8·5	1·2	2·12	20	2·38
Magnesium	450	7·7	1·1	1·92	19	2·66
Nickel	655	7·7	1·4	2·80	30	2·47
Copper	450	7·5	0·9	1·88	20	nil to 0·34
Iron	450	7·7	1·2	1·92	20	5·62*

* Hydrogen peroxide equivalent to 18·4 mg iron.

The results show that zinc, aluminium, magnesium and nickel produce approximately 3 mg H_2O_2 in one hour. Copper and iron behave differently; copper gives very little or no hydrogen peroxide, whilst iron gives an amount of ferric ions greater than the equivalent of 3 mg H_2O_2.

With all the metals the water becomes turbid during the experiment indicating that substantial quantities of metal hydroxides had become suspended in the solution. The experimental set-up was unsuitable for collecting the hydroxides quantitatively. In one experiment with zinc the equivalent of 11·4 mg Zn $(OH)_2$ was found to be present.

2.4. *H_2O_2 Concentration as a Function of Number of Cuts and of Time*

Experiments were carried out with zinc to determine the concentration of hydrogen peroxide after a number of cuts had been made. The conditions of the experiments were the same as reported in §2.3 and the number of cuts was varied. The estimation of the concentration of hydrogen peroxide was carried out immediately after cutting. The results of these experiments are shown in fig. 1. After an initial rapid rise, the concentration of hydrogen peroxide increases linearly with the number of cuts, i.e. with the increase in surface area exposed by cutting. The significance of the shape of the curve will be discussed later. It is interesting to note at this stage that the use of a saturated solution of zinc oxide reduces the amount of hydrogen peroxide formed as can be seen from table 2.

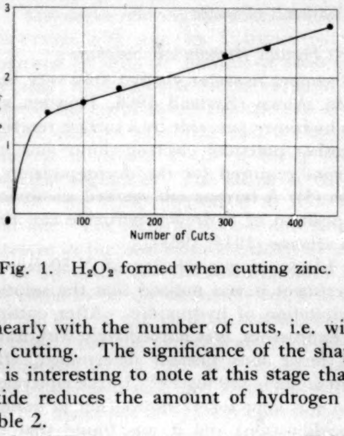


Fig. 1. H_2O_2 formed when cutting zinc.

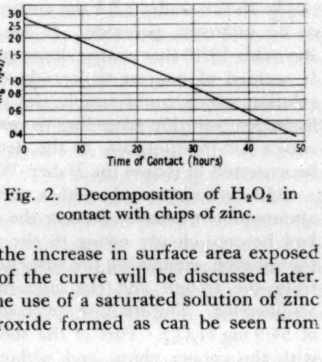


Fig. 2. Decomposition of H_2O_2 in contact with chips of zinc.

Table 2

No. of cuts	50	100
Cut in water (mg H_2O_2)	1·46	1·53
Cut in water saturated with ZnO (mg H_2O_2)	0·85	1·02

In several instances 200 ml. samples of water were withdrawn from the trough after a number of cuts had been made, and replaced by 200 ml. of fresh water. Cutting was continued to a total number of 450 cuts. The amount of hydrogen peroxide present was then found to be greater than if cutting was continued without exchanging part of the water. If the exchange of 200 ml. of water was carried out after 150 cuts the final concentration was 3·49 mg H_2O_2/l. as against 2·8 mg H_2O_2/l. with no exchange. If exchanged after 300 cuts the final concentration was 3·92 mg H_2O_2/l.

Hydrogen peroxide solutions, obtained from the cutting of zinc, were found to be relatively stable, if kept separate from the metal chips. When kept in contact with the chips the concentration of hydrogen peroxide decreased. Experiments were carried out to determine the rate of decomposition of hydrogen peroxide when in contact with chips of zinc. In every case 450 cuts were made, as described in § 2.3. The solutions were left in contact with the chips for definite periods of time and the trough was kept covered with an opaque plate in order to prevent photolysis of hydrogen peroxide. Figure 2 shows the change in concentration of hydrogen peroxide with time.

2.5. *Cutting of Zinc in Air*

It seemed of interest to see whether the formation of hydrogen peroxide occurred during the time of actual cutting or as an after effect of this process. A zinc block was fixed to the trough and 400 ml. of distilled water was added. This quantity of water was insufficient to cover the block and cutting could thus proceed in air and the chips immediately immersed in water. Only chips from 350 cuts could be accommodated in the volume of water available. The quantity of hydrogen peroxide developed in the water was 1·52 mg as compared with 2·45 mg when cutting 350 chips under 1000 ml. of water.

2.6. *Cutting of Copper in N/50 Ferrous Ammonium Sulphate*

As shown under § 2.3 the cutting of copper in water yielded only very little or no hydrogen peroxide. It has been shown (Saeland 1908, Dowden and Reynolds 1950) that copper decomposes hydrogen peroxide by a surface reaction. It seemed of interest to introduce another potential electron donor into the solution which could supply the electrons required for the decomposition of hydrogen peroxide according to reaction (5). A ferrous salt seemed an obvious choice for this purpose, as the decomposition of hydrogen peroxide can then be expected to follow the Haber–Weiss scheme (1934, 1935).

450 cuts were made with a copper block using a solution of N/50 ferrous ammonium sulphate. During the experiment it was noticed that the solution had become cloudy owing to the precipitation of hydroxides. After cutting, the solution, containing the suspended hydroxides, was immediately withdrawn from the trough and the amount of ferric ions present determined, after acidification. The solution was found to contain 180 mg Fe^{3+}/l. (the equivalent of 54·9 mg H_2O_2). Part of the solution was kept for 17 hours (not in contact with the copper chips, and without acidification) and it was found that the concentration of ferric ions had increased to 300 mg (the equivalent of 93 mg H_2O_2). Part of the latter solution was used for a copper estimation. The hydroxides were filtered off and estimations were carried out both on the filtrate and in the precipitate. The filtrate was found to contain the equivalent of 4 mg copper/l. and the precipitate the equivalent of about 50 mg copper/l.

§ 3. DISCUSSION

The experiments show that when fresh metallic surfaces are produced under water, by the cutting of zinc, aluminium, magnesium and nickel, measurable quantities of hydrogen peroxide are formed. Under similar conditions the metals mentioned produce approximately the same quantity of hydrogen peroxide. With copper only very small quantities of hydrogen peroxide can occasionally be detected. This behaviour is similar to that observed in the Russell effect. It therefore appears that phenomena similar to the Russell effect can be produced by a process of 'mechanical activation'.

Some peculiar property of fresh metallic surfaces is involved and there seems little doubt that some form of electron transfer is responsible for the reactions. It is yet uncertain whether the transfer of electrons occurs as a homogeneous or as a heterogeneous process. The same uncertainty exists about the Kramer effect which is likely to be one of the sources of the electrons.

The transfer of an electron is probably facilitated by a partially ionic character of a bond in a molecule. In the case of a water molecule, having the configuration $H^{+}\,{}^{-}OH$, the electron attaches itself to the proton and causes the ionic field to disappear. The result is a hydrogen atom and a hydroxyl ion. The free hydrogen atom can initiate various free radical reactions (Waters 1946) and in the presence of oxygen and water leads to the formation of hydrogen peroxide and hydroxyl radicals, .OH. The latter can also form hydrogen peroxide by the reaction $2.OH \rightarrow H_2O_2$. This reaction requires a three-body collision in order to dispose of the energy of the free radicals and should occur rather infrequently.

The curve relating the concentration of hydrogen peroxide to the number of cuts made with zinc (fig. 1) shows that a very rapid rise occurs at first, which is followed by a linear increase of concentration with the increase in the number of cuts. Leaving aside for the moment the initial rise, the linearity of the greater part of the curve shows that the rate of increase in concentration is proportional to the increase in surface area. This observation implies that within the time interval in which a cut is being made, the 'activity' of the surface has become negligible. The shape of the linear part of the curve indicates that 3×10^{-3} mg H_2O_2 (equivalent to $5{\cdot}3 \times 10^{16}$ molecules) are being formed during one cut. The latter figure can now be compared with the number of zinc atoms present in the new surface created by a cut. The surface density of zinc atoms is of the order of $1{\cdot}6 \times 10^{15}$ atoms/cm^2. The surface area created in a cut can be taken to be approximately equal to twice the area laid bare, if the reduction in the size of the chips is considered to compensate for the roughness of the surface. For the experiments with zinc this would correspond to 22·1 cm^2/cut, equivalent to $3{\cdot}3 \times 10^{16}$ zinc atoms. It thus appears that in the later stages of the experiment each zinc atom exposed is roughly responsible for the formation of 1·5 molecules of hydrogen peroxide.

The above considerations cannot be applied to the initial rise in concentration. It seems obvious that this rise cannot be caused by some property of the metal, since it is reasonable to assume that zinc atoms behave similarly, whether exposed in the early or in the later stages of an experiment. One must therefore assume that a change in the properties of the solution is responsible. Various hypotheses can be advanced to explain the rapid initial rise in concentration.

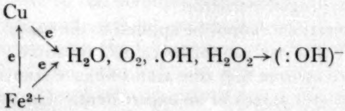
(i) The water is rapidly depleted of oxygen. The amount of oxygen present (21 mg) is, however, more than sufficient for the reaction to proceed at a uniform rate. Experiments with water containing 36 mg of oxygen did not yield results materially different from those reported. This hypothesis can therefore be dismissed.

(ii) Saturation is reached with respect to one of the products of the reaction. This suggestion may imply reversibility, not normally observed with electron transfer reactions. Saturation can, however, be involved in a different way. The solution, after the experiment, contained a considerable quantity of zinc hydroxide in suspension. It seems plausible that before the solution had become saturated with Zn $(OH)_2$, a considerably greater area of metal had been laid bare by the dissolution of Zn $(OH)_2$ formed on the surface than by the process of cutting. The intercept of the curve in fig. 1 corresponds to about 1·3 mg H_2O_2. If it is assumed that about 1·5 molecules of hydrogen peroxide are formed per zinc atom exposed, the dissolution process can be considered to be responsible for the exposure of 2×10^{19} zinc atoms or the equivalent amount of surface exposed in about 55 cuts. Again from the curve it is evident that saturation is reached after approximately 50 cuts. The rough correspondence between the number of cuts at which saturation is reached and the surface area required to account for the hydrogen peroxide formed in the initial stages, suggests that at first the layer of zinc hydroxide is removed by dissolution exposing a fresh metal surface. When the solution has become saturated with respect to Zn $(OH)_2$ the dissolution process ceases and only one surface layer of zinc atoms is allowed to react. Although this hypothesis is supported by the results given in table 2, it does not fully account for the relatively large amounts of H_2O_2 formed when, during an experiment, part of the solution is replaced by fresh distilled water.

When the solution of hydrogen peroxide is left in contact with zinc metal chips the logarithm of the concentration falls off approximately linearly with time, indicating that the reaction was pseudo-unimolecular in character, with an apparent half-time of about 17 hours.

The results obtained with iron require some comment. As with other metals a suspension of hydroxides was obtained, part of which was ferric hydroxide. The amount of ferric hydroxide obtained was only slightly greater than the equivalent amount of hydrogen peroxide from other metals.

Copper in the presence of a ferrous salt gave rise to 180 mg ferric ions, the equivalent of 54·9 mg H_2O_2. The oxygen content of the water was 22·1 mg equivalent of 47 mg H_2O_2. There seems to be a rough correspondence between the amount of oxygen present and the amount of ferric ions produced, which allows the conclusion that in this case the extent of the reaction is limited by the amount of oxygen available. The difference between the two equivalents can be attributed to replenishment of oxygen during the course of the experiment. The transfer of electrons probably follows the scheme



The oxidation of ferrous ions proceeds even after the removal of the copper chips. A sufficient amount of copper is being retained by the solution and by the hydroxide precipitate for the above reaction scheme to continue.

ACKNOWLEDGMENTS

The work described in this paper was carried out in the Lubrication Division of the Mechanical Engineering Research Laboratory, Department of Scientific and Industrial Research, as part of an item on the programme of the Mechanical

Engineering Research Board. The paper is published by permission of the Director of the Laboratory.

The author wishes to acknowledge the assistance of Dr. K. H. R. Wright and Mr. J. A. Houston.

REFERENCES

CHURCHILL, J. R., 1939, *Trans. Electrochem. Soc.*, **76,** 341.

DOWDEN, D. A., and REYNOLDS, P. W., 1950, *Discussions of the Faraday Society*, **8,** 184.

DUNSTAN, W. R., JOWETT, H. A. D., and GOULDING, E., 1905, *J. Chem. Soc.*, **87,** 1548.

FIANDA, F., and LANGE, E. L., 1951, *Z. Electrochem.*, **55,** 237.

GRUNBERG, L., and WRIGHT, K. H. R., 1952, *Nature, Lond.*, **170,** 456.

HABER, F., and WEISS, J., 1934, *Proc. Roy. Soc.* A, **147,** 332 ; 1935, *Trans. Faraday Soc.*, **31,** 1547.

HAXEL, O., HOUTERMAN, F. G., and SEEGER, K., 1951, *Z. Phys.*, **130,** 109.

KEENAN, G. L., 1920, *Chem. Rev.*, **3,** 95.

KRAMER, J., 1949, *Z. Phys.*, **125,** 739; 1950 a, *Der metallische Zustand* (Göttingen : Vandemhoeck and Ruprecht); 1950 b, *Z. Phys.*, **128,** 538; 1951, *Ibid.*, **129,** 34.

LOUW, J. D., and NAUDÉ, S. M., 1949, *Phys. Rev.*, **76,** 571.

RUSSELL, W. J., 1897, *Proc. Roy. Soc.*, **61,** 424; 1898, *Ibid.*, **63,** 102; 1899, *Ibid.*, **64,** 409; 1906, *Ibid.*, **78,** 385; 1908, *Ibid.*, **80,** 376.

SAELAND, S., 1908, *Ann. Phys., Lpz.*, **26,** 899.

SHAW, M. C., 1948, *J. Appl. Mech.*, **15,** 37.

WATERS, W. A., 1946, *The Chemistry of Free Radicals* (Oxford : Clarendon Press).

Long Wavelength Infra Red Photoconductivity of Silicon at Low Temperatures

BY B. V. ROLLIN AND E. L. SIMMONS
The Clarendon Laboratory, Oxford

MS. received 17*th November* 1952

Abstract. The conductivity of silicon at low temperatures is found to be increased by illumination with infra-red radiation in the wavelength range 2 to 14 μ. The effect is believed to be due to excitation of electrons or holes from impurity states. At 14°K, for radiation of wavelength 10 μ, interrupted at 800 c/s, a signal equal to noise in a bandwidth of 1 c/s is obtained with an incident intensity of about 2×10^{-8} watt/cm^2 on a specimen of area 0·1 cm^2.

§1. INTRODUCTION

THE energy gap between the highest filled band in silicon and the conduction band is known to be about 1·1 ev (Torrey and Whitmer 1948) and intrinsic photoconductivity, due to transitions between these bands, is observed for radiation of wavelength about 1 μ. The object of the present work was to investigate photoconductivity due to impurity ionization which should occur at much longer wavelengths.

The energy states of an electron or hole bound in a hydrogen-like orbit to an isolated impurity centre in a medium of dielectric constant K are given by $E_n = 13{\cdot}5/K^2n^2$ ev. For silicon, $K = 13$ so that the ionization energy is approximately 0·08 ev. The ionization energy is reduced by the presence of neighbouring impurities (Pearson and Bardeen 1949) and it is found that, for p type silicon containing N_A acceptors per cm^3:

$$E_I = (0{\cdot}08 - 4{\cdot}13 \times 10^{-8} N_A{}^{1/3})\ \text{ev}.$$

For a specimen of moderate purity with $N_A = 10^{17}/\text{cm}^3$, $E_I = 0{\cdot}06$ ev, so that photoconductivity due to impurity ionization should be possible for wavelengths less than 20 μ.

The photoconductivity can only be observed if the rate of ionization by radiation is comparable with the rate of thermal ionization, which is proportional to $\exp(-E_I/\boldsymbol{k}T)$. With such a small ionization energy, it is necessary to make measurements at low temperatures. For $E_I = 0{\cdot}06$ ev temperatures of the order of 20°K are required.

In order to have a reasonable value of E_I it is desirable to use a specimen with a fairly small impurity concentration but, on the other hand, an appreciable fraction of the incident radiation must be utilized in impurity ionization. The cross section for impurity ionization in silicon is estimated to be of the order of 10^{-15} cm^2 near the ionization limit and it should diminish as $\lambda^{3{\cdot}5}$ for wavelengths much shorter than this (Hall 1936, Burstein *et al.* 1951) so that, for an impurity concentration of about $10^{17}/\text{cm}^3$, a specimen of thickness of the order of 1 mm is required. The lattice absorption in silicon for long wavelength radiation is very small (Briggs 1950, Lord 1952) and, for a specimen of this thickness, is only appreciable in the neighbourhood of absorption bands at 9 μ and 16·5 μ. At low temperatures, absorption due to mobile electrons or holes can be neglected.

§ 2. EXPERIMENTAL ARRANGEMENT

In one form of apparatus the specimen was situated in a Dewar vessel containing liquid hydrogen or liquid helium and radiation from a rock salt prism monochromator was directed on to it through a K.R.S. 5 window at the top of the apparatus. Measurements were normally made with the level of the liquid below the specimen, otherwise trouble was experienced due to scattering of the radiation by bubbles in the liquid. The surfaces of the specimen were polished and contacts soldered to nickel plated areas on the ends. The specimen was mounted below a slit in a radiation shield and was connected by a short, screened cable to a preamplifier which was followed by a tuned amplifier with a 4 c/s bandwidth. The radiation was interrupted by a sector disc at 20, 360 or 800 c/s. The resistance of the specimen at 20°K was very high (of the order of 100 to 1000 MΩ) and it was necessary to increase the impedance of the preamplifier and the input cable by the use of negative feedback.

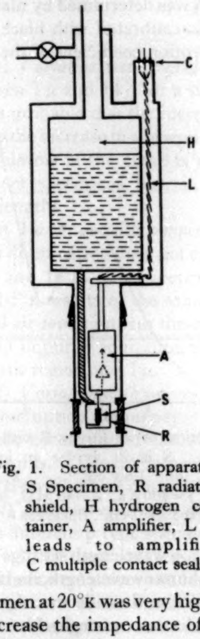


Fig. 1. Section of apparatus. S Specimen, R radiation shield, H hydrogen container, A amplifier, L six leads to amplifier, C multiple contact seal.

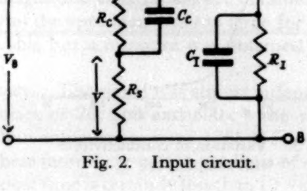


Fig. 2. Input circuit.

An alternative form of apparatus is shown in fig. 1. The specimen S was situated in the same vacuum enclosure as the liquid hydrogen container H and was cooled by conduction along one of the leads. The specimen was surrounded by a radiation shield R which was joined to the hydrogen container by a copper rod. The first stage of the preamplifier A was also in the vacuum enclosure and as near to the specimen as possible so as to reduce the capacity of the input lead. By operating the first valve as a cathode follower, a very high input impedance could be obtained. Radiation from the monochromator passed through a window in the wall of the vacuum enclosure and through a slit in the radiation shield.

In the input circuit, shown in fig. 2, R_s represents the bulk resistance of the specimen, R_c and C_c are the contact resistance and capacity and R_I and C_I are the input resistance and capacity of the preamplifier and cable. The feed resistor R_F was of about the same magnitude as R_s. The impedance of the coupling condenser C and the contact capacity C_c at 20 c/s was negligible. For a voltage V across R_s the signal ΔV at AB due to a change of conductance ΔG of the specimen is given by

$$\Delta V = VZ\Delta G \qquad \ldots\ldots(1)$$

where Z is the impedance between AB:

$$\frac{1}{Z} = \frac{1}{R_s} + \frac{1}{R_F} + \frac{1}{R_I} + j\omega C_I. \qquad \ldots\ldots(2)$$

The intensity of radiation falling on the specimen was determined by measurements made with a thermistor bolometer which was calibrated with black body radiation from an oven at 350°C, assuming the absorption coefficient of the bolometer to be independent of wavelength.

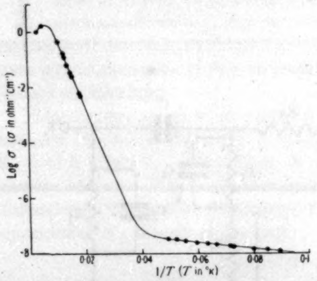


Fig. 3. Variation of conductivity σ with temperature T.

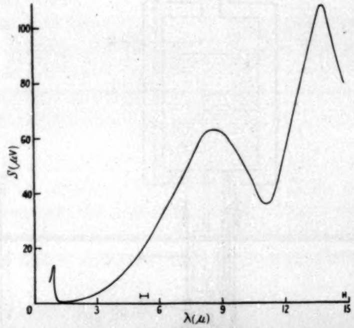


Fig. 4. Variation of sensitivity S with wavelength λ. S is in μv for an incident intensity of $1\ \mu$w cm^{-2}. $T = 20$°K, $V = 40$ v, $f = 800$ c/s, $Z = 40\ \mathrm{M}\Omega$, area of specimen $0{\cdot}5 \times 0{\cdot}2$ cm^2, thickness 0·16 cm.

In order to make sure that the signal observed at long wavelength settings of the monochromator was not due to scattered radiation of shorter wavelength, the change of signal on inserting filters of lithium fluoride or calcium fluoride was observed. It was found that, in all cases, the effect of scattered radiation was negligible.

§ 3. RESULTS

(i) *Variation of Conductivity with Temperature*

The change of conductivity with temperature for a single crystal p type specimen is shown in fig. 3. Between 90°K and 50°K the relation $\sigma = \sigma_0 \exp(-E_I/\boldsymbol{k}T)$ is valid with $E_I = 0{\cdot}05$ ev and $\sigma_0 = 250\ \mathrm{ohm}^{-1}\,\mathrm{cm}^{-1}$. At lower temperatures, however, the conductivity is considerably greater than that given by this expression. This result may be due to the existence of a wide distribution of impurity levels or, alternatively, to conduction by another mechanism, such as injection of minority carriers. Similar results obtained with germanium (Gerritsen 1949, Hung and Gliessman 1950) have been attributed to conduction in the impurity band (Hung 1950).

(ii) *Absorption Coefficient*

At room temperature the fraction of incident radiation transmitted by a polycrystalline specimen of thickness 1 mm was 50% at 2 μ and 44% at 10 μ. At 20°K the transmission was 50% at 2 μ but only 30% at 10 μ. The absorption coefficient at 20°K was therefore about 2 cm^{-1} at 10 μ. Since this specimen contained about 2×10^{16} impurities per cm^3 the cross section for impurity ionization at 10 μ is about 10^{-16} cm^2.

(iii) *Photoconductivity*

(*a*) *Variation with wavelength*. The variation of response with wavelength between 1 μ and 14 μ for a single crystal specimen at 20°K is shown in fig. 4. It seems probable that the increase in response with increasing wavelength is mainly due to the increase in absorption in the specimen. The explanation for the maxima and minima in the curve is not known. The shape of the curve is different for different specimens but, in all cases, there is a general rise towards longer wavelengths.

(*b*) *Variation with temperature*. The change of conductance produced by a given illumination does not change appreciably in the temperature region between 20°K and 14°K. In experiments at liquid helium temperatures between 4·2°K and 1·5°K results of the same order of magnitude as at 20°K were obtained. At liquid air temperatures the conductivity of the specimen was too large for effects due to impurity ionization to be observable but a response was obtained in the intrinsic region near 1 μ.

(*c*) *Variation with interruption frequency*. The signal was almost independent of interruption frequency. At frequencies of 20, 360 and 800 c/s the relative signals, corrected for the change of input impedance, were 176, 155 and 158 microvolts. The differences between these figures are within the limit of experimental error and we can say that the response time is certainly less than 10^{-4} second.

(*d*) *Different specimens and heat treatment*. Measurements have been made on three different p type specimens with impurity content between 2×10^{16} and 10^{17} per cm^3. Of these, one was the single crystal for which results have been given and the others were polycrystalline but containing only a few large crystals. The order of magnitude of the signal at 20°K was the same for all specimens. An improvement in the signal to noise ratio by a factor of about two was obtained by annealing the single crystal at 450°C for 12 hours.

(*e*) *Signal to noise ratio*. The noise voltage at the input of the amplifier was due to the combined effects of Johnson noise from the specimen, shot and flicker noise in the first valve and current noise in the specimen. In order to obtain optimum signal to noise ratio the voltage across the specimen was increased until the current noise was the predominant factor. Since this type of noise has a power spectrum inversely proportional to frequency (Montgomery 1952) the optimum signal to noise ratio was obtained at the highest interruption frequency. For a given voltage across the specimen the noise voltage diminished with decreasing temperature (see table) so that the signal to noise ratio was best at the lowest temperatures.

Temperature (°K)	20	20	14	14
Frequency (c/s)	20	800	20	800
Voltage across specimen (v)	90	90	108	108
Current (μA)	0·43	0·43	0·18	0·18
Noise per unit bandwidth (μv r.m.s.)	18	2·5	9	1·7

Input impedance $Z=40$ MΩ Noise with $V=0$: 2·5 μv at 20 c/s; 0·85 μv at 800 c/s

For the single crystal specimen of area 0·1 cm^2 at 14°K the minimum detectable intensity of radiation at 10 μ for an interruption frequency of 800 c/s and a bandwidth of 1 c/s was about 2×10^{-8} watt cm^{-2}.

§4. DISCUSSION

(i) *Bolometer Effect*

We must first consider whether the long wavelength response could be due to a bolometer effect.

Suppose that the increase of temperature ΔT of the specimen due to heat input W is $\Delta T = W/k$ so that the change of conductance is $\Delta G = (dG/dT)W/k$. It is difficult to estimate k directly but it can be expressed in terms of the response time $\tau = C/k$ where C is the heat capacity of the specimen so that

$$\Delta G = (\tau W/C)\, dG/dT.$$

The change of conductance due to radiation of intensity I falling on a specimen of area a and absorption coefficient α is therefore $\Delta G = (\tau\alpha a I/C)\, dG/dT$ and the resulting signal is given by $\Delta V = VZ\Delta G = (VZ\tau\alpha a I/C)\, dG/dT$. Inserting numerical values for a single crystal specimen at 20°K with specific heat of 0·1 joule mol^{-1} deg^{-1} (Keesom and Pearlman 1952) and assuming $\tau = 10^{-4}$ sec and $\alpha = 1$ we obtain $\Delta V \simeq 10^{-1}\,\mu$v for $I = 1\,\mu$w cm^{-2}, $V = 100$ v, $Z = 4 \times 10^7$ ohm. This is about a thousand times smaller than the observed signal at long wavelengths.

(ii) *Impurity Ionization*

We now have to consider whether the results are consistent with those to be expected for impurity ionization.

Suppose we have a specimen containing N_{D} donors and N_{A} acceptors per cm^3 with N_{D} greater than N_{A} so that the specimen is n type. If n donors per cm^3 are thermally ionized at temperature T there are: n conduction electrons, $(N_{\mathrm{D}} - N_{\mathrm{A}} - n)$ neutral donors, $(N_{\mathrm{A}} + n)$ ionized donors and N_{A} ionized acceptors.* If A is the probability per unit time of an electron combining with an ionized donor and B is the probability of thermal ionization of a neutral donor then, at equilibrium:

$$An(N_{\mathrm{A}} + n) = B(N_{\mathrm{D}} - N_{\mathrm{A}} - n). \qquad \ldots\ldots(3)$$

At very low temperatures n will be negligible in comparison with N_{A} so that

$$nN_{\mathrm{A}}/(N_{\mathrm{D}} - N_{\mathrm{A}}) = B/A = K. \qquad \ldots\ldots(4)$$

The equilibrium constant K is given by (Fowler 1936):

$$K = 2(2\pi \boldsymbol{m}\boldsymbol{k}T/\boldsymbol{h}^2)^{3/2} \exp(-E_{\mathrm{I}}/\boldsymbol{k}T) \simeq 5 \times 10^{15} T^{3/2} \exp(-E_{\mathrm{I}}/\boldsymbol{k}T). \qquad \ldots\ldots(5)$$

---

* When current is flowing through the specimen this equilibrium condition may be disturbed by holes injected from the positive contact. At low temperatures the holes will be trapped by ionized acceptors so that there may be a reduction in the number of ionized impurities.

Suppose that, with radiation of intensity I, Q quanta are absorbed per cm^3 per second and are effective in producing ionization. The change in the number of conduction electrons Δn is given by:

$$A\Delta n N_{\mathrm{A}} = Q, \qquad \ldots\ldots(6)$$

so that, if β is the mobility, the change of conductivity $\Delta\sigma$ is:

$$\Delta\sigma = \boldsymbol{e}\beta Q/N_{\mathrm{A}}A. \qquad \ldots\ldots(7\,a)$$

If τ is the relaxation time for capture of a conduction electron by an ionized impurity

$$\tau = 1/N_{\mathrm{A}}A, \qquad \ldots\ldots(8\,a)$$

so that

$$\Delta\sigma = \boldsymbol{e}\beta Q\tau. \qquad \ldots\ldots(9)$$

For a p type specimen the equations are identical except that N_{D} and N_{A} are interchanged so that, instead of equations (7 *a*) and (8 *a*) we have

$$\Delta\sigma = \boldsymbol{e}\beta Q/N_{\mathrm{D}}A, \qquad \ldots\ldots(7\,b)$$

$$\tau = 1/N_{\mathrm{D}}A. \qquad \ldots\ldots(8\,b)$$

For the single crystal specimen at 20°K with $I = 1\,\mu$w cm^{-2} at $10\,\mu$, corresponding to $Q \simeq 5\times10^{13}$ cm^{-3} sec^{-1}, $\Delta\sigma = 4\times10^{-13}$ ohm^{-1} cm^{-1}. Assuming $\beta \simeq 1000$ cm^2 sec^{-1} v^{-1}, we obtain from eqn. (9), $\tau \simeq 5\times10^{-11}$ second. There is no direct experimental evidence that this value of τ is correct but it appears to be somewhere near the value to be expected. If v is the average thermal velocity of a mobile hole and σ_{h} is the cross section for capture by an ionized acceptor we have

$$\tau = (N_{\mathrm{D}}\sigma_{\mathrm{h}}v)^{-1}. \qquad \ldots\ldots(10)$$

Unfortunately we do not know the values of either σ_{h} or N_{D} and we can only make a very rough guess as to their order of magnitude. It seems probable that N_{D} is likely to be of the order of a tenth of the total number of acceptors or about 10^{16}/cm^3. For σ_{h} we assume a value of about 10^{-14} cm^2 which is the order of magnitude found for capture of electrons by F centres in the alkali halides (Mott and Gurney 1948). With $v = (\boldsymbol{k}T/2\boldsymbol{m})^{1/2}$ the value obtained for τ at 20°K is of the order of 10^{-9} second.

Another estimate of τ can be obtained from B by means of equations (5) and (8). It is found that the rectification properties of silicon at microwave frequencies can be understood if, at room temperatures, B has a value of about 10^{10} sec^{-1} (Torrey and Whitmer 1948). If we assume $B = B_0 \exp(-E_{\mathrm{I}}/\boldsymbol{k}T)$, with $B_0 \simeq 10^{11}$ sec^{-1} this leads to the values $B \simeq 10^{-1}$ sec^{-1}, $A \simeq 10^{8}$ sec^{-1} cm^3 at 20°K so that, from eqn. (8), with $N_{\mathrm{D}} = 10^{16}$/cm^3 we obtain $\tau \simeq 10^{-8}$ second. Obviously these estimates of τ are very crude and all that can be said is that they indicate a value in the region of $10^{-9\pm1}$ second.

Attempts have been made to evaluate B by quantum mechanical perturbation calculations (Goodman, Lawson and Schiff 1947, Kubo 1952) but the validity of the assumptions appears to be in some doubt and no reliance can safely be placed on the results.

(iii) *Sensitivity and Response Time*

In view of the possibility of silicon being useful as a detector of long wavelength infra-red radiation, it is interesting to consider whether there is any chance of increasing the sensitivity above the present level. It is apparent from eqns. (1) and (9) that increase of sensitivity can only be obtained at the expense of increase of response time but, if the response time is indeed as short as 10^{-10} sec, this would be no disadvantage since it is impossible to make effective use of a response time shorter than about 10^{-5} sec on account of the difficulty of obtaining a sufficiently high input impedance of the amplifier at high frequencies. Since the response time depends on the concentration of the minority type of impurity there is a possibility of gaining a factor of the order of ten by careful control of the purity of the specimen. The response is also dependent on the magnitude of A. At the present time very little is known about the factors which influence the transfer of energy between the electrons and the lattice vibrations but it seems probable that the magnitude of A and B_0 may depend on whether or not the ionization energy is less than that corresponding to the highest lattice vibration frequency. The value of B_0 for silicon appears indeed to be appreciably higher than for impurity activated phosphors for which it is about $10^{8\pm1}$ sec^{-1} (Garlick 1949). If this is the case, other substances with a suitable ionization energy might be more sensitive than silicon as detectors of long wave infra-red radiation.

ACKNOWLEDGMENTS

We are greatly indebted to Mr. R. W. Douglas of the Research Laboratories of the General Electric Company for supplying us with some of the specimens of silicon. This work has been supported by a grant from the Admiralty and we wish to thank the Director of Physical Research for permission to publish these results. One of us (E.L.S.) gratefully acknowledges the receipt of a maintenance grant from the Department of Scientific and Industrial Research.

REFERENCES

BRIGGS, H. B., 1950, *Phys. Rev.*, **77**, 727.
BURSTEIN, E., OBERLY, J. J., DAVISSON, J. W., and HENVIS, B. W., 1951, *Phys. Rev.*, **82**, 764 (A).
FOWLER, R. H., 1936, *Statistical Mechanics* (Cambridge : University Press).
GARLICK, G. F. J., 1949, *Rep. Prog. Phys.*, **12**, 34 (London : Physical Society).
GERRITSEN, A. N., 1949, *Physica*, **15**, 427.
GOODMAN, B., LAWSON, A. W., SCHIFF, L. I., 1947, *Phys. Rev.*, **71**, 191.
HALL, H., 1936, *Rev. Mod. Phys.*, **8**, 358.
HUNG, C. S., 1950, *Phys. Rev.*, **79**, 727.
HUNG, C. S., and GLIESSMAN, J. R., 1950, *Phys. Rev.*, **79**, 726.
KEESOM, P. H., and PEARLMAN, N., 1952, *Phys. Rev.*, **85**, 730 (A).
KUBO, R., 1952, *Phys. Rev.*, **86**, 929.
LORD, R. C., 1952, *Phys. Rev.*, **85**, 140.
MONTGOMERY, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.
MOTT, N. F., and GURNEY, R. W., 1948, *Electronic Processes in Ionic Crystals* (Oxford : University Press).
PEARSON, G. L., and BARDEEN, J., 1949, *Phys. Rev.*, **75**, 865.
TORREY, H. C., and WHITMER, C. A., 1948, *Crystal Rectifiers* (New York : McGraw-Hill).

The Energy Balance Equation for the Positive Columns of High Pressure Arcs

By K. S. W. CHAMPION*
Electron Physics Department, University of Birmingham

Communicated by J. Sayers; MS. received 21st November 1952

Abstract. Expressions for the energy balance equation put forward by various authors are reviewed and the need for a more rigorous statement of this equation is pointed out. An equation is deduced which includes rigorous expressions for the conduction and convection losses and an approximate expression for the radiation loss. Another equation which must be solved simultaneously is the hydrodynamical equation of steady flow, which yields the gas drift velocity involved in the convection losses. These general equations, unfortunately, can only be solved for certain special sets of conditions.

A solution is given for a special case of a cylindrically symmetrical vertical arc and it is found that on insertion of appropriate numerical values the calculated quantities are in good agreement with experiment.

---

§ 1. Introduction

So far no completely rigorous energy balance equation has been put forward for high pressure arcs from which energy is lost by conduction, convection and radiation. The well-known Elenbaas–Heller relation only represents the conduction loss. Suits and Poritsky (1939) have shown that the conduction and convection losses from high pressure arcs can be represented with reasonable accuracy by the Nusselt relation, which holds for the losses from a hot solid cylinder in a gas. Some refinements in their analysis, due to Champion (1952), result in better agreement between theory and experiment. The convection losses from a vertical arc have also been considered by Elenbaas (1951) who, in addition, evaluated graphically the vertical component of the convection velocity for one particular mercury arc. The difficult problem of representing the radiation losses from arcs has been attacked by several writers, including Francis (1946, 1949).

Thus it can be seen that all the present statements of the energy balance equation have some limitations. Consequently, the possibility of deriving a more rigorous expression for the energy balance equation has been investigated.

§ 2. The General Energy Balance Equation

The Elenbaas–Heller equation

$$-\frac{1}{r}\frac{d}{dr}\left(rk\frac{dT}{dr}\right)=e\,n_e KE^2 \qquad \ldots\ldots(1)$$

---

* Now at Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

relates the energy dissipated in a volume element of an arc column to the energy lost by conduction in a radial direction. In this expression E is the positive column gradient, n_e the electron density and K the total electron and ion mobility and it is assumed that the temperature T is a function only of the radial coordinate r. The general expression for the conduction loss when T may be a function of all the coordinates is $-\operatorname{div}(k \operatorname{grad} T)$.

In deriving the expression for the convection loss it will be assumed that the arc is in a steady state, that is the physical properties (e.g. gas velocity and temperature) at any point are independent of time. From consideration of the losses from an element of the arc the convection loss is found to be $\operatorname{div}(sT\rho\mathbf{v})$ where ρ, s and $\mathbf{v}$ are the gas density, specific heat and drift velocity respectively. It is not possible to give a rigorous expression for the radiation losses. However, the expression $Bn_e\exp(-\boldsymbol{e}V_{ex}/\boldsymbol{k}T)$ where V_{ex} is the excitation potential given by Koch, Lesemann and Walther (1949) is quite useful, although the value of the radiation constant B will depend markedly on the gas and other conditions and will not always be accurately constant throughout a particular discharge. Thus the general expression for the energy balance equation is

$$\operatorname{div}(sT\rho\mathbf{v}) - \operatorname{div}(k \operatorname{grad} T) + Bn_e \exp(-\boldsymbol{e}V_{ex}/\boldsymbol{k}T) = \boldsymbol{e}n_eKE^2 \qquad \ldots\ldots(2)$$

The first term of eqn. (2) can be expressed in an alternative form by making use of the equation of continuity for the steady state, viz : $\operatorname{div}(\rho\mathbf{v}) = 0$. Thus

$$\operatorname{div}(sT\rho\mathbf{v}) = sT\operatorname{div}(\rho\mathbf{v}) + \rho\mathbf{v}\operatorname{grad}(sT) = \rho\mathbf{v}\operatorname{grad}(sT) \qquad \ldots\ldots(3)$$

and

$$\rho\mathbf{v}\operatorname{grad}(sT) - \operatorname{div}(k \operatorname{grad} T) + Bn_e \exp(-\boldsymbol{e}V_{ex}/\boldsymbol{k}T) = \boldsymbol{e}n_eKE^2 \qquad \ldots\ldots(4)$$

The terms in eqn. (4) representing the energy dissipation and conduction and convection losses are completely rigorous and the term representing the radiation loss is a good approximation. However, the solution of this equation for the general case is very difficult, despite the fact that outside the ionized region of the arc two of the terms are zero. The difficulty is not greatly reduced for those arcs (such as most arcs in permanent gases) for which the radiation loss from the body of the arc is sufficiently small to be neglected.

The specific heat of the gas s can generally be represented by c_p since, with the arc in a steady state (dynamic equilibrium), the variation in pressure between the various parts of the discharge tube is usually small. With dissociating gases c_p is a function of temperature and the effect of ionization (e.g. the thermal type given by Saha's equation) causes an additional temperature variation in c_p.

§ 3. THE GAS DRIFT VELOCITY

Another quantity whose value is required before eqn. (4) can be solved is the gas drift velocity $\mathbf{v}$. This is determined by the hydrodynamical equation of steady, viscous flow which is

$$(\operatorname{div} \eta \operatorname{grad})\mathbf{v} = \operatorname{grad} p - \rho\mathbf{g}, \qquad \ldots\ldots(5)$$

where η is the viscosity coefficient.

In cases where the radiation losses can be neglected eqns. (4) and (5) together determine rigorously the convection and conduction losses from a high pressure arc. Their solution also yields such associated properties as the gas temperature distribution and convection velocity. However, it seems that in practice their solution will be restricted to a few special sets of conditions.

For a cylindrically symmetrical vertical arc eqn. (5) may be expressed in the form

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\eta\frac{\partial v_z}{\partial r}\right)+\frac{\partial}{\partial z}\left(\eta\frac{\partial v_z}{\partial z}\right)=\frac{\partial p}{\partial z}+\rho g \qquad \ldots\ldots(6)$$

and

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\eta\frac{\partial v_r}{\partial r}\right)+\frac{\partial}{\partial z}\left(\eta\frac{\partial v_r}{\partial z}\right)=\frac{\partial p}{\partial r} \qquad \ldots\ldots(7)$$

where the z axis coincides with the axis of the arc. A simple case of these conditions occurs when v_z is independent of z and $\partial p/\partial z$ is constant. Then

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\eta\frac{\partial v_z}{\partial r}\right)=(\rho-\rho_c)g \qquad \ldots\ldots(8)$$

where $\rho_c g$ is a constant. Now $\rho=\rho_c T_c/T$ where T_c is the gas temperature corresponding to density ρ_c, but the functional variation of T depends on the as yet unsolved equation (4). However, with the temperature distributions found experimentally in arcs $T_{\mathrm{arc}}\{1+\xi(r/a)^{\varphi}\}^{-1}$ provides a good representation for T. ξ and ϕ are constants determined by the individual conditions and T_{arc} is the temperature at the axis of the arc. Thus

$$\frac{\partial}{\partial r}\left(r\eta\frac{\partial v_z}{\partial r}\right)=\rho_{\mathrm{arc}}g\,r\left\{1-\frac{T_{\mathrm{arc}}}{T_c}+\xi\left(\frac{r}{a}\right)^{\varphi}\right\}$$

and

$$\eta\frac{\partial v_z}{\partial r}=\rho_{\mathrm{arc}}g\left[\left(1-\frac{T_{\mathrm{arc}}}{T_c}\right)\frac{r}{2}+\frac{\xi r^{\varphi+1}}{(\phi+2)\,a^{\varphi}}\right]$$

where the integration constant is zero since $(\partial v_z/\partial r)_{r=0}=0$.

Now $\eta=\eta_{\mathrm{arc}}(T/T_{\mathrm{arc}})^{\theta}=\eta_{\mathrm{arc}}\{1+\xi(r/a)^{\varphi}\}^{-\theta}$ where the exponent θ governs the temperature variation of the viscosity. Theoretically $\theta=0{\cdot}5$ for a rigid elastic sphere model of a neutral gas and is greater for all other models. Experimentally observed values of θ range between 0·65 and 1·1, but there are no results for gases at very high temperatures or for those containing a proportion of ionized atoms. It would be expected that, in general, the presence of thermal ionization would increase the value of θ.

The solution for v_z when $\theta=1$ is straightforward.

$$v_z=\frac{\rho_{\mathrm{arc}}g}{\eta_{\mathrm{arc}}}\left[\frac{\xi^2 r^{2\varphi+2}}{2(\phi+1)(\phi+2)a^{2\varphi}}+\left\{\frac{1}{\phi+2}+\frac{1}{2}\left(1-\frac{T_{\mathrm{arc}}}{T_c}\right)\right\}\frac{\xi r^{\varphi+2}}{(\phi+2)a^{\varphi}}+\left(1-\frac{T_{\mathrm{arc}}}{T_c}\right)\frac{r^2}{4}\right]+\text{const.} \qquad \ldots\ldots(9)$$

The value of the constant can immediately be obtained from the boundary condition $(v_z)_{r=a}=0$. The boundary condition

$$\int_0^a r\rho\,v_z\,dr=0$$

which expresses the fact that the same mass of gas rises as falls due to convection, gives the value of T_c. On reduction this relation becomes

$$-\frac{1}{16}-\frac{\xi}{4(\phi+4)}+\frac{\phi+2\xi+2}{8(\phi+2)}+\frac{\xi(\phi+2\xi+2)}{4(\phi+2)^2}-\frac{\xi^2}{2(\phi+2)^3}+\frac{\xi^2(\phi\xi+2\xi+2\phi+2)}{2(\phi+1)(\phi+2)^3}$$
$$-\frac{\xi^3}{2(\phi+1)(\phi+2)(3\phi+4)}=\frac{T_{\mathrm{arc}}}{4T_c}\left\{\frac{1}{4}+\frac{\xi(\phi+\xi+2)}{(\phi+2)^2}\right\}. \qquad \ldots\ldots(10)$$

Resort must be made to series to obtain the general solution for v_z for any value of θ. Two expansions are required for $\{1+\xi(r/a)^{\varphi}\}^{\theta}$ to cover the regions in which

$\xi(r/a)^{\varphi}$ is greater and less than unity respectively. The integral for the former region yields

$$v_z = \frac{\rho_{\mathrm{arc}} g}{\eta_{\mathrm{arc}}} \left\{ \left(1 - \frac{T_{\mathrm{arc}}}{T_{\mathrm{c}}}\right) \Sigma \frac{\chi_i \xi^{\theta-i} r^{\varphi\theta - i\varphi + 2}}{2(\phi\theta - i\phi + 2)\, a^{\varphi\theta - i\varphi}} + \frac{1}{\phi+2} \Sigma \frac{\chi_i \xi^{\theta-i+1} r^{\varphi\theta+\varphi-i\varphi+2}}{(\phi\theta+\phi-i\phi+2)\, a^{\varphi\theta+\varphi-i\varphi}} \right\} + \text{const.} \qquad \ldots\ldots(11)^*$$

where, as before, the value of the constant is determined by the condition $(v_z)_{r=a} = 0$. The integral for the region in which $\xi(r/a)^{\varphi} < 1$ gives

$$v_z = \frac{\rho_{\mathrm{arc}} g}{\eta_{\mathrm{arc}}} \left\{ \left(1 - \frac{T_{\mathrm{arc}}}{T_{\mathrm{c}}}\right) \Sigma \frac{\chi_i \xi^{i} r^{i\varphi+2}}{2(i\phi+2) a^{i\varphi}} + \frac{1}{\phi+2} \Sigma \frac{\chi_i \xi^{i+1} r^{\varphi+i\varphi+2}}{(\phi+i\phi+2) a^{\varphi+i\varphi}} \right\} + \text{const.} \qquad \ldots\ldots(12)$$

Both series for v_z converge when $r = a/\xi^{1/\varphi}$ and the value of the constant in (12) is obtained by equating the two expressions at this point. T_{c} is evaluated as before and the appropriate relation is

$$\begin{aligned}
&\frac{\xi^{\theta}}{4} \Sigma \frac{\chi_i \xi^{-i}}{\phi\theta - i\phi + 4} + \frac{\xi^{\theta+1}}{\phi+2} \Sigma \frac{\chi_i \xi^{-i}}{\phi\theta - i\phi + \phi + 4} + \frac{\xi^{\theta+2}}{(\phi+2)^2} \Sigma \frac{\chi_i \xi^{-i}}{\phi\theta - i\phi + 2\phi + 4} \\
&\quad + \frac{1}{4\xi^{4/\varphi}} \Sigma \left(\frac{\chi_i}{i\phi+4} - \frac{\chi_i}{\phi\theta - i\phi + 4} \right) + \frac{1}{(\phi+2)\,\xi^{4/\varphi}} \Sigma \left(\frac{\chi_i}{i\phi+\phi+4} - \frac{\chi_i}{\phi\theta - i\phi + \phi + 4} \right) \\
&\quad + \frac{1}{(\phi+2)^2 \xi^{4/\varphi}} \Sigma \left(\frac{\chi_i}{i\phi+2\phi+4} - \frac{\chi_i}{\phi\theta - i\phi + 2\phi + 4} \right) = \frac{T_{\mathrm{arc}}}{T_{\mathrm{c}}} \left\{ \frac{\xi^{\theta}}{4} \Sigma \frac{\chi_i \xi^{-i}}{\phi\theta - i\phi + 4} \right. \\
&\quad + \frac{\xi^{\theta+1}}{2(\phi+2)} \Sigma \frac{\chi_i \xi^{-i}}{\phi\theta - i\phi + \phi + 4} + \frac{1}{4\xi^{4/\varphi}} \Sigma \left(\frac{\chi_i}{i\phi+4} - \frac{\chi_i}{\phi\theta - i\phi + 4} \right) \\
&\quad \left. + \frac{1}{2(\phi+2)\xi^{4/\varphi}} \Sigma \left(\frac{\chi_i}{i\phi+\phi+4} - \frac{\chi_i}{\phi\theta - i\phi + \phi + 4} \right) \right\} \qquad \ldots\ldots(13)
\end{aligned}$$

§ 4. Comparison with Experiment

The expression $T_{\mathrm{arc}}\{1 + \xi(r/a)^{\varphi}\}^{-1}$ fits most accurately the experimental temperature distributions found in arcs in diatomic and polyatomic gases, as in these gases the temperature is nearly constant in the body of the arc and falls sharply at the boundaries. In fig. 1 are plotted two typical curves (*c* and *d*) for T which could fit such conditions. However, the fit is still good in the case of arcs in monatomic gases. This can be illustrated by comparison with the temperature distribution given by Elenbaas (1951) for his vertical arc in mercury vapour which he calls 'discharge D'. In fig. 1 are plotted Elenbaas' curve for T (curve *a*) and also the expression $5885\,\{1 + 4{\cdot}5(r/a)^4\}^{-1}$ (curve *b*). The fit is relatively good but, in any case, there will be a little inaccuracy in Elenbaas' curve as the greater part of it is obtained by calculation assuming that there is no transfer of energy by convection.

The curve for v_z was calculated from eqns. (11) and (12) with the above expression for T and the value 0·75 for θ. This curve is plotted in fig. 2 (curve *d*) and the close agreement with Elenbaas' curve ($\theta = 0{\cdot}75$) (curve *a*) is apparent. However, the latter was obtained by graphical integration and involved solution of an equation containing a triple integral. This equation could only be solved by trial and error.

---

* χ_i represents the factors 1, θ, $\frac{\theta(\theta-1)}{2}$, $\frac{\theta(\theta-1)(\theta-2)}{3!}$ ... for successive values 0, 1, 2, 3 ... of i. Σ throughout this paper denotes $\sum\limits_{i=0}^{\infty}$.

To illustrate how v_z depends on the value of θ the curves for $\theta=0$ and 1 have been calculated and plotted in fig. 2 (curves *b* and *c*). Although, with present knowledge, there can be no certainty as to the value of θ applicable to an arc it seems that the value 1 may be closer to the actual value than 0·75 for the

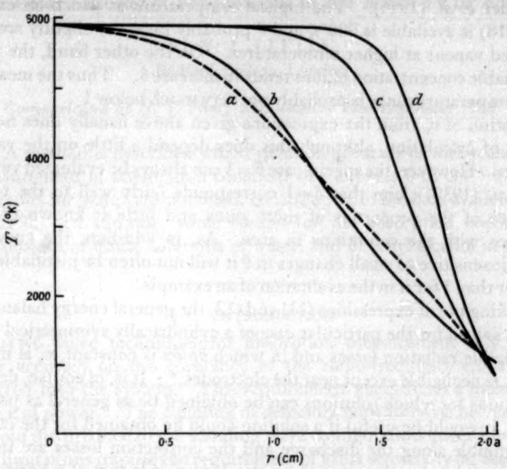


Fig. 1. *a*. Elenbaas' temperature distribution for his discharge D. *b*. Curve calculated for this discharge from author's expression for temperature. *c, d*. Typical curves for the temperature distributions in arcs in polyatomic gases.

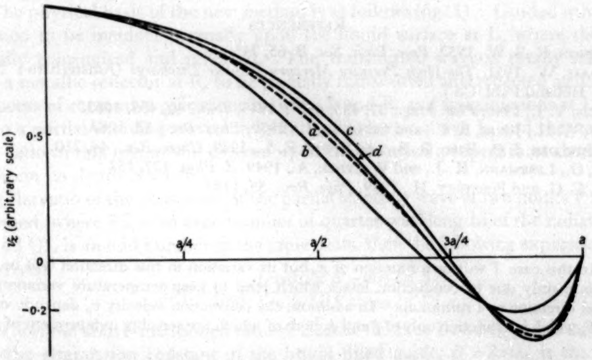


Fig. 2. *a*. Vertical convection velocity obtained graphically by Elenbaas for his discharge. *b, c, d*. Calculated velocity for this discharge assuming that $\theta=0, 1, 0·75$ respectively. For ease of comparison an arbitrary scale is used showing V_z as unity when $r=0$ in each case.

'discharge D'. The experimental and theoretical values given for mercury vapour by Elenbaas (1951, p. 42) support this value as do also the values quoted by Hirschfelder *et al.* (1949). The highest temperature at which an experimental value (1·018) is available is 900°K and θ probably becomes slightly smaller in the non-ionized vapour at higher temperatures. On the other hand, the presence of an appreciable concentration of ions tends to increase θ. Thus the mean value of θ over the temperature range is probably not very much below 1.

Evaluation of v_z from the expressions given above usually does not involve a great deal of calculation, although this does depend a little on the values of the parameters. However, the special case $\theta = 1$ can always be evaluated very speedily. Furry *et al.* (1939) claim that $\theta = 1$ corresponds fairly well to the temperature dependence of the properties of most gases and little is known of the actual dependence with the conditions in arcs. As, in addition, the curve for v_z is relatively insensitive to small changes in θ it will not often be justifiable to use any value other than 1 for it in the evaluation of an example.

By making use of expressions (11) and (12) the general energy balance equation (4) can be solved for the particular case of a cylindrically symmetrical vertical arc with negligible radiation losses and in which $\partial p/\partial z$ is constant, v_z is independent of z and v_r is negligible except near the electrodes.* It is, of course, desirable that the conditions for which solutions can be obtained be as general as possible. In particular, it would be useful if a solution could be obtained for the case when v_r is not negligible along the discharge and the convection losses are increased by 'eddy currents'. Another case of practical interest is that of a horizontal arc.

While these solutions are awaited the expression and analysis of the energy balance equation given by Champion (1952) will be found to be of considerable value.

REFERENCES

CHAMPION, K. S. W., 1952, *Proc. Phys. Soc.* B, **65**, 345.

ELENBAAS, W., 1951, *The High Pressure Mercury Vapour Discharge* (Amsterdam: North-Holland Publ. Co.).

FRANCIS, V. J., 1946, *Phil. Mag.*, **37**, 433, 653; 1949, *Ibid.*, **40**, 435, 1063.

FURRY, W. H., JONES, R. C., and ONSAGER, L., 1939, *Phys. Rev.*, **55**, 1083.

HIRSCHFELDER, J. O., BIRD, R. B., and SPOTZ, E. L., 1949, *Chem. Rev.*, **44**, 210.

KOCH, O., LESEMANN, K.-J., and WALTHER, A., 1949, *Z. Phys.* **127**, 153.

SUITS, C. G. and PORITSKY, H., 1939, *Phys. Rev.*, **55**, 1184.

---

* In this case T will be a function of z, but its variation in this direction will be small. This is mainly due to conduction losses which tend to keep temperature variations in a vertical direction to a minimum. In addition, the convection velocity v_z depends only on $T_{\mathrm{arc}}/T_{\mathrm{c}}$ which is a function only of ξ and ϕ, both of which are sensibly independent of z.

A Standing Wave Method for Measuring Electromagnetic Absorption in Polar Liquids at Frequencies of the order 3×10^9 c/s

By V. I. LITTLE
Bedford College, London

Communicated by H. T. Flint; MS. received 22nd September 1952

Abstract. A method is described which yields an accuracy of within about ±1% in the measurement of attenuation constants of the order $0{\cdot}01\ \mathrm{cm}^{-1}$ to $0{\cdot}5\ \mathrm{cm}^{-1}$. Accurate results for water are obtained, contributing to the data available for this liquid at 21°c. In addition, some results for the absorption coefficient and dielectric constant of dilute solutions of water, and of ethyl alcohol in dioxan are provided.

---

§ 1. INTRODUCTION

STANDING wave techniques for absorption measurement rely partly for their accuracy on the constancy of the radio-frequency e.m.f. which is injected into the system. This requirement necessitates a loose coupling to the source of power. The coupling of detectors must likewise be very weak, in order to avoid disturbance to the standing wave configuration under investigation. These two limitations impose the requirement of great sensitivity on the detectors, which is usually accomplished by the use of calibrated heterodyne receivers.

If two simple detectors are employed, however, it becomes possible to couple the oscillator critically to the standing wave apparatus, with the consequent advantages accompanying an increase of available signal strength.

The physical basis of the new method is as follows (fig. 1) : Guided waves are assumed to be incident normally upon the liquid surface at L, where they are partially transmitted and reflected. The transmitted wave is totally reflected from a metallic reflector at R, to be partially transmitted and reflected again at L. A process of successive reflections between L and R, and transmissions at L gives rise to a partial standing wave configuration in the air above L, and the standing wave ratio in this region will depend upon the electrical properties of the liquid, and upon its depth $S = \mathrm{LR}$.

If the ratio of the amplitude of the partial standing wave at two points P and Q is formed, where PL is an even number of quarter wavelengths of the radiation in air, and QL is an odd number of the same units, then the following expression can be derived

$$|\rho|^2 = |\rho_0|^2 \left\{ \frac{\cosh 2\alpha' S - \cos 2\beta' S}{\cosh 2\alpha' S + \cos 2\beta' S} \right\} \qquad \ldots\ldots(1)$$

where ρ is the above-mentioned ratio of electric fields at P and Q, ρ_0 is a constant, α' is the attenuation constant of the liquid filled guide, $\beta' = 2\pi/\lambda_{\mathrm{L}}$ is the phase constant of the liquid filled guide with λ_{L} the wavelength of the radiation in the liquid.

§ 2. DERIVATION OF EQUATION (1)

Flint and Pincherle (1943) have shown that in waveguide propagation it is possible to set up potential functions F and U, which behave analogously to potential difference and current in normal twin conductor transmission systems.

Thus, for H waves in wave guides, a stream function **F** is defined, such that the electric field E is given by $E = -\text{curl}\ \mathbf{F}$, where **F** lies in the direction of propagation and F is its amplitude.

Similarly U is defined by equating the transverse component of the magnetic field in the guide to the negative gradient of a scalar U.

With these substitutions the following equations are derived

$$F = F_0 \cosh Pz - Z_0 U_0 \sinh Pz \qquad \ldots\ldots(2)$$

and

$$U = U_0 \cosh Pz - (F_0/Z_0) \sinh Pz \qquad \ldots\ldots(3)$$

where Z_0 and P are the effective characteristic impedance and propagation constant of the guide respectively.

These equations are similar to the standard transmission line equations describing the voltage and current at any point on a transmission line terminated in a general impedance, in terms of the voltage and current at any chosen point on the line.

It follows therefore in fig. 1 that

$$F_0'/U_0' = Z_0' \tanh P'S \qquad \ldots\ldots(4)$$

where F_0' and U_0' are the values of F and U at the liquid–air interface, and Z_0' is the characteristic impedance of the liquid filled portion of the guide.

In the region above the liquid surface, the values of F and U at any point, l cm above the interface are given by

$$F = F_0 \cosh P(L-l) - Z_0 U_0 \sinh P(L-l) \qquad \ldots\ldots(5)$$

and

$$U = U_0 \cosh P(L-l) - (F_0/Z_0) \sinh P(L-l) \qquad \ldots\ldots(6)$$

where F_0 and U_0 are the values of F and U at any arbitrary point a distance L above the interface, and Z_0 is the characteristic impedance of the air-filled portion of the guide.

When $l = 0$, the boundary conditions are $F = F_0'$ and $U = U_0'$ at the interface. Using this condition, it follows that

$$Z_0' \tanh P'S = \frac{F_0 \cosh PL - Z_0 U_0 \sinh PL}{U_0 \cosh PL - (F_0/Z_0) \sinh PL}$$

or

$$\frac{F_0}{U_0} = \frac{Z_0' \tanh P'S \cosh PL + Z_0 \sinh PL}{\cosh PL + (Z_0'/Z_0) \sinh PL \tanh P'S}. \qquad \ldots\ldots(7)$$

From equations (5) and (7) it follows that

$$\frac{F}{F_0} = \frac{Z_0' \tanh P'S \cosh Pl + Z_0 \sinh Pl}{Z_0' \tanh P'S \cosh PL + Z_0 \sinh PL}. \qquad \ldots\ldots(8)$$

Further, if F_1 and F_2 are two values of F, at distances l_1 and l_2 respectively from the interface, the ratio F_1/F_2 may be formed, whence

$$\frac{F_1}{F_2} = \frac{Z_0' \tanh P'S \cosh Pl_1 + Z_0 \sinh Pl_1}{Z_0' \tanh P'S \cosh Pl_2 + Z_0 \sinh Pl_2}. \qquad \ldots\ldots(9)$$

In the air filled portion of the guide, $P = j\beta = 2\pi j/\lambda$ where λ is the wavelength of the radiation in this portion of the guide. If this substitution is made in (9) and $l_1 = \frac{1}{2}\lambda$, and $l_2 = \frac{3}{4}\lambda$,

$$\frac{F_1}{F_2} \text{ reduces to } -j\frac{Z_0'}{Z_0} \tanh P'S. \qquad \ldots\ldots(10)$$

Now F_1 and F_2 are quantities which are linearly proportional to the electric field in the waveguide, and square-law detectors coupled to the field at these points will give signals proportional to the squares of the moduli of these quantities. It follows that the ratio of the signals from two square-law detectors coupled to the guide at l_1 and l_2 will give

$$\left|\frac{F_1}{F_2}\right|^2 = |\rho|^2 = \left|\frac{Z_0'}{Z_0}\right|^2 |\tanh P'S|^2. \qquad \ldots\ldots(11)$$

Writing, $P' = \alpha' + j\beta'$, and $|\rho_0|^2 = |Z_0'/Z_0|^2$, and substituting in eqn. (11),

$$|\rho|^2 = |\rho_0|^2 \tanh(\alpha' + j\beta')S \tanh(\alpha' - j\beta')S$$

which reduces to

$$|\rho|^2 = |\rho_0|^2 \frac{\cosh 2\alpha' S - \cos 2\beta' S}{\cosh 2\alpha' S + \cos 2\beta' S},$$

which is the required equation.

For the purposes of experiment, a graph is plotted of $|\rho|^2$ against S. Such a graph as plotted from eqn. (1) will be found in fig. 2.

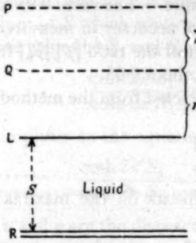


Fig. 1.

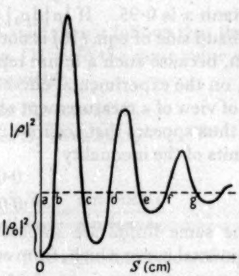


Fig. 2. Theoretical curve for liquid of high attenuation and high dielectric constant.

§ 3. INTERPRETATION OF EXPERIMENTAL CURVE

Equation (1) indicates that $|\rho|^2$ has the value $|\rho_0|^2$, whenever $\cos 2\beta' S = 0$, i.e. when $S = (\frac{1}{8} + \frac{1}{4}n)\lambda_L$. Points where this condition is met are shown in fig. 2 at b, c, d, e, f, etc. The intervals bc, cd, ef, etc. are all equal to $\lambda_L/4$ and the most accurate way of finding λ_L consists of drawing a line through the experimental curve parallel to the axis of S, so that the curve cuts the line into equal portions. Such a line may be adjusted with considerable accuracy (within ± 1%), and once fixed the value of $|\rho_0|^2$ is automatically given, and the experimental curve effectively normalized.

The most direct method of evaluating α' is to measure the value of $|\rho|^2$ at points such that $\cos 2\beta' S = \pm 1$.

Equation (1) reduces to

$$|\rho|^2 = |\rho_0|^2 \tanh^2 \alpha' S \quad \text{when} \quad \cos 2\beta' S = +1, \qquad \ldots\ldots(12)$$

or

$$|\rho|^2 = |\rho_0|^2 \coth^2 \alpha' S \quad \text{when} \quad \cos 2\beta' S = -1. \qquad \ldots\ldots(13)$$

$$\cos 2\beta' S = +1 \quad \text{when} \quad \beta' S = n\pi \quad \text{or} \quad S = \tfrac{1}{2} n \lambda_L, \qquad \ldots\ldots(14)$$

$$\cos 2\beta' S = -1 \quad \text{when} \quad \beta' S = \tfrac{1}{2}\pi[2n+1] \quad \text{or} \quad S = \tfrac{1}{4}\lambda_L[2n+1]. \qquad \ldots\ldots(15)$$

Condition (14) and eqn. (12) refer to points near the minima of the curve of eqn. (1), whilst condition (15) and eqn. (13) fix points near the maxima. Equation (12) defines a curve tangential to eqn. (1) at the minima, whilst eqn. (13) defines a curve tangential to equation (1) at the maxima. The points of contact of eqn. (12) with eqn. (1) are at values of S given by condition (14), i.e. at points midway between cd, ef, gh, etc. Likewise the points of contact of eqn. (13) with eqn. (1) are at values of S given by condition (15), i.e. at points midway between bc, de, fg, etc.

These points are thus very easily found, once the line abcd etc. has been fixed, and the value of the ordinates of the experimental curve at these points when substituted into either eqn. (12) or (13) (whichever applies) yields α' very quickly.

Let us consider the practical limitations associated with the extraction of α' from the experimental curve. We note that eqn. (12) applies when $S = \frac{1}{2}n\lambda_L$ where n is an integer. Substituting this value for S, eqn. (12) becomes

$$\tanh^{-1}(|\rho|^2/|\rho_0|^2)^{1/2} = \tfrac{1}{2}\alpha' n\lambda_L. \qquad \ldots\ldots(16)$$

The positive square root has been taken, since α' is essentially positive. Now $\tanh x$ varies from 0 to unity as x varies from 0 to infinity. Furthermore, when x is 1·83, $\tanh x$ is 0·95. If $|\rho|/|\rho_0|$ is much in excess of 0·95, then the value of the right-hand side of eqn. (16) is not accurately calculable. The value 0·95 has been chosen, because such a figure represents an order of accuracy in measurement of $\pm 2\%$ on the experimental curve. The lower limit of the ratio $|\rho|/|\rho_0|$ from the point of view of a measurement accurate to $\pm 2\%$ is about 0·05.

It thus appears that accurate results may be extracted from the method within the limits of the inequality

$$0{\cdot}05 \leqslant |\rho|/|\rho_0| \leqslant 0{\cdot}95$$

or

$$0{\cdot}05 \leqslant \tfrac{1}{2}\alpha n\lambda_L \leqslant 1{\cdot}83. \qquad \ldots\ldots(17)$$

The same limits are imposed upon measurements on the maxima of the experimental curve which, from eqn. (13) yields

$$0{\cdot}05 \leqslant \tfrac{1}{2}\alpha(n+\tfrac{1}{2})\lambda_L \leqslant 1{\cdot}83. \qquad \ldots\ldots(17\,a)$$

The inequalities (17) and (17 *a*) indicate that values of α' and λ_L which satisfy them for a series of values of n, may be obtained with considerable accuracy. This is true to a certain extent, but another practical limitation soon becomes evident.

Consider the case of a liquid whose attenuation constant is 0·001 cm^{-1} and whose dielectric constant is 2. The effective wavelength of 10 cm radiation in such a liquid will be of the order 7 cm. Using equation (17) we can calculate the minimum value of $\frac{1}{2}n\lambda_L$ which will allow the attenuation constant to be measured. This is evidently $0{\cdot}05/0{\cdot}001 = 50$ cm. A liquid depth of at least this amount must be used. Clearly, for such a liquid, the method would be too cumbersome.

If a practical maximum depth of liquid of the order 10 cm is adopted, then for liquids which satisfy the relationship $\frac{1}{2}n\lambda_L = 10$, the minimum value of the attenuation constant which falls within the measurable range of the method is 0·005 cm^{-1}.

The case for high absorption is rather different. The inequality (17) gives $\frac{1}{2}\alpha' n\lambda_L = 1{\cdot}8$, or $\alpha' = 3{\cdot}6/n\lambda_L$. Taking the limiting case in which n is unity, then the maximum value of α' which can be measured is $3{\cdot}6/\lambda_L$. Evidently the smaller λ_L the more useful does the method become.

As an illustration it will suffice to compare two liquids such as water and ethyl alcohol. The attenuation constants of these liquids are roughly the same, and lie in the neighbourhood of 0·5 cm^{-1} for radiation of wavelength 10 cm in air. The wavelength of such radiation in the liquids, however, is markedly different. For water λ_L is about 1 cm and for alcohol about 4 cm.

Substitution of these values into the inequality (17) yields a value for n in both cases—i.e. the number of minima in the experimental curve. For water $n=7$, whereas for alcohol $n=2$.

The factor which determines λ_L is the real part of the dielectric constant. For water this is 80, and for alcohol about 4 or 5. Clearly the method will yield best results for aqueous solutions in which the dielectric constant is still very high, for under these circumstances the experimental curve will have a considerable number of maxima and minima, all of which will lie in the measurable region, yet confined within some 10 cm of liquid depth. Highly absorbing liquids like ethyl alcohol, however, will not yield good results because of their relatively low dielectric constants at frequencies of the order 3×10^9 c/s.

For highly absorbing liquids with low dielectric constants, the experimental curve may be interpreted in the following manner. In such cases only the first maximum is obtained, the subsequent minimum being too small to detect. The wavelength λ_L is still measurable, though with less accuracy, by finding a line such as abc (fig. 3) which runs parallel to the S axis, and is intersected by the curve in such a way that bc = 2 ab.

If another line XYZ is taken parallel to the S axis, then it follows that there are two values of S which when substituted into eqn. (1) yield the same value of $|\rho|^2$. Let these values be S_1 and S_2. Equation (1) then yields

$$\frac{\cosh 2\alpha' S_1-\cos 2\beta' S_1}{\cosh 2\alpha' S_1+\cos 2\beta' S_1}=\frac{\cosh 2\alpha' S_2-\cos 2\beta' S_2}{\cosh 2\alpha' S_2+\cos 2\beta' S_2}. \qquad \ldots\ldots(18)$$

This simplifies to the equation

$$\frac{\cosh 2\alpha' S_1}{\cosh 2\alpha' S_2}=\frac{\cos 2\beta' S_1}{\cos 2\beta' S_2}=\frac{\cos 4\pi S_1/\lambda_L}{\cos 4\pi S_2/\lambda_L}=\frac{\sin 4\pi u/\lambda_L}{\sin 4\pi v/\lambda_L} \qquad \ldots\ldots(19)$$

where u and v are the distances shown in fig. 3. The right-hand side of eqn. (19) can be evaluated, leaving α to be calculated by a process of successive approximations. Many lines such as XYZ may be drawn, each yielding a value of α from which a mean value may be extracted. Much labour, however, is required in the arithmetical solution of eqn. (19), and to shorten the time involved for this process, an analogue computer was developed. This computer will be described in a subsequent paper (Little 1953).

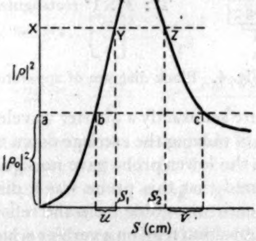


Fig. 3. Theoretical curve for liquid of high attenuation and low dielectric constant.

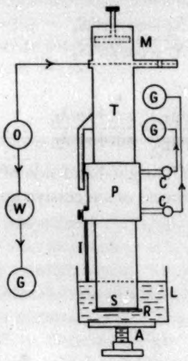
A third method of solution is also available. If the slope of the experimental curve is measured at the points b, c, d, etc. then by differentiation of eqn. (1) with respect to S, and insertion of the condition that $\cos 2\beta' S = 0$, it follows that

$$\left|\frac{\text{slope of curve at b}}{\text{slope of curve at c}}\right| = \frac{\cosh 2\alpha'(\text{ac})}{\cosh 2\alpha'(\text{ab})}. \qquad \ldots\ldots(20)$$

This equation can also be solved arithmetically, or with the aid of the computer.

§ 4. Apparatus and Experimental Procedure

The results which follow were obtained with a guided wave system consisting of a length of 3 in. × 1 in. rectangular waveguide about a metre long, terminated by means of a metal plate. The H_{01} mode of propagation was employed, and the liquid was allowed to flow into the guide through two narrow slots cut in the narrow faces of the guide flush with the terminating plate. The liquid was contained in a cylindrical glass vessel, sufficiently wide to accommodate the cross section of the guide. A narrow slot cut down the centre of one broad face of the guide facilitated the measurement of the standing wave amplitude in the guide by means of two crystal probes, mounted together on a sliding carriage. Signals from a coaxial type oscillator were fed into the guide at the upper end of the apparatus, and into an N.P.L. type of cavity wavemeter.



Block Diagram of Apparatus

A : adjustable table
C : crystal probe detectors
G : sensitive galvanometers
I : index marker
L : liquid container
M : matching section
O : oscillator
P : probe carriage
R : reflecting plate, terminating guide
S : slot cut in narrow face of guide at termination to allow liquid access
T : 3″ × 1″ rectangular waveguide

Fig. 4. Block diagram of apparatus.

Initially the probes were set exactly a quarter wavelength apart by an electrical method. This consisted of moving the carriage down towards the short circuit to the lowest point at which the lower probe gave no signal in the absence of liquid in the guide. This ensured that this probe was a distance equal to one half a wavelength of the radiation in the guide from the reflector. The probe carriage was then clamped, and its position read on a vernier scale clamped to the apparatus mounting.

The upper probe, which was adjustable in position on the probe carriage, was then moved until it recorded a maximum signal. An index marker, consisting of a length of brass rod, was then adjusted so that its point lay in the same plane as the reflector plate, outside the guide. This adjustment was performed with the aid of a travelling microscope. Once adjusted under these conditions, the upper end of the index marker was clamped to the probe carriage.

The liquid reservoir was then raised up so that the horizontally adjusted reflecting surface became immersed in the liquid, and the latter allowed to flow inside the guide. The probe carriage was raised, and lowered gradually, until the index marker just broke the surface of the liquid outside the guide. This adjustment could be made with great consistency, the error in positioning the probes with respect to the liquid surface being less than 0·01 cm.

A series of readings was taken in this way, of the detector signals for various depths of liquid, the latter being read off directly on the vernier scale on the carriage, or by means of the travelling microscope previously used for adjusting the index marker.

One very important aspect of the method must be mentioned at this stage. The coupling of the probes to the waveguide must be very weak, to avoid interactions between them, and to prevent any disturbance of the standing wave configuration within the guide. The twin probe method is admirably suited to this requirement, because the apparatus may be critically coupled to the source of oscillations, and provided that both crystals obey the same power law, amplitude variations of the source do not affect the ratio of the signal amplitudes at the detectors.

Figure 4 is a diagrammatical representation of the apparatus. Two crystal detectors were found in a batch of four dozen, to have the same law, and high sensitivity. In order to check that the laws of the crystals were the same, it was only necessary to adjust the probe carriage so that various signal ratios were obtained and to note that the same ratios were maintained for wide variations of input power. With this point satisfactorily checked, the input coupling control was used to keep one of the probe signals at a constant level during the course of an experiment in which case the other probe signal could be taken as being directly proportional to the ratio of the two signals.

In this way it was possible to plot the whole of an experimental curve directly —a matter of two to three hundred points—in about one hour. The wavemeter served to monitor the frequency throughout the experiment.

The above procedure provides experimental data from which the values of α' and β' for the liquid filled portion of the waveguide can be calculated.

These quantities are then related to the electrical properties of the liquid by the equations

$$\alpha'^2-\beta'^2=\left[\frac{\pi}{a}\right]^2-\frac{\mu\epsilon'\omega^2}{\mathbf{c}^2},$$

$$2\alpha'\beta'=\frac{\omega^2\mu\epsilon''}{\mathbf{c}^2} \quad \text{or} \quad \frac{\omega\mu 4\pi\gamma}{\mathbf{c}},$$

where a is the broad dimension of the waveguide supporting the H_{01} mode, ϵ' and ϵ'' the real and imaginary parts of the dielectric constant of the liquid, γ the conductivity of the liquid in e.s.u., ω the angular frequency of the supply and $\mathbf{c}$ the velocity of light. These equations are standard results and may be derived from data provided by Montgomery (1947, Chap. 10).

In these equations everything is known except ϵ', ϵ'' and γ and these may thus be calculated.

§ 5. RESULTS

Using the above method, results for water were obtained for radiations whose wavelength in air were 11·12, 10·57, 10, 9·75 and 9·16 cm. These values represented the limits of operation of the signal generator employed. They are usefully grouped, however, to exhibit the consistency of working of the apparatus.

In table 1, values of α', λ_L and derived values of ϵ' and ϵ'' at 21°c are shown. In this table α_1' is the value obtained by the method of minima or maxima measurement, α_2' is obtained by measurement on the slopes of the curve, and α_3' is obtained by measurements on the intercepts of a line drawn on the experimental plot, parallel to the depth axis. Estimated accuracies are indicated, each value of α' being a mean of several determinations.

Table 1. Values obtained for the Absorption Coefficient of Water at 21°c

λ_{air} (cm)	λ_L (cm)	ϵ	ϵ'	α_1'	α_2'	α_3'	$\gamma \times 10^{-10}$ (e.s.u.)
9·16	1·040	77·0	14·8	0·58 ± 2%	0·59 ± 2%	0·56 ± 5%	2·1
9·75	1·105	77·6	14·0	0·50 ± 2%	0·51 ± 2%	0·53 ± 5%	2·2
10·00	1·130	78·0	13·5	0·48 ± 1%	0·48 ± 2%	0·51 ± 5%	2·0
10·57	1·195	78·1	13·0	0·44 ± 1%	0·44 ± 1%	0·46 ± 5%	1·8
11·12	1·254	78·5	12·2	0·39 ± 1%	0·39 ± 1%	0·40 ± 5%	1·6

When dealing with absorption in polar liquids, it is convenient to plot ϵ'' against ϵ'. The resulting locus, according to Debye's theory of relaxation absorption (1929), should be a semicircle for liquids with single relaxation times, if measurements are made throughout the absorbing region. This method of presentation of results was first demonstrated by Cole and Cole (1941). The above results, for water at 21°c, are displayed by this method in fig. 5, together with results obtained recently at 20°c, by Collie, Ritson and Hasted (1948), Cook (1951), and by Saxton and Lane (1952). The results of these three groups of workers are undoubtedly the most accurate yet published, and were obtained by the method of waveguide attenuation in guides of different dimensions, and by resonator methods. The value for the high-frequency dielectric constant of water has been taken as 5, following Saxton (1952), whilst the low frequency value at 20° c given by Collie *et al.* (1948) is taken as 80·36.

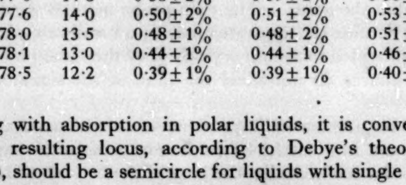


Fig. 5. Cole and Cole plot for water at 20° c. Results of recent work.

Key to wavelengths (cm) for points plotted in fig. 5.

	1	2	3	4	5		
A	1·27	3·2	10·00	—	—	Collie *et al.* 1948	20°c
B	0·62	1·24	3·21	—	—	Saxton and Lane 1952	
C	1·262	3·195	6·84	10·00	—	Cook 1951	
D	9·16	9·75	10·00	10·57	11·12	Author	21°c

It will be observed that a significant difference exists between the 10 cm result of the present method and the results of Cook and of Collie *et al.* This difference is probably due to the effect of the liquid meniscus in the standing wave method. At the frequencies employed, errors due to this effect could not be gauged with certainty, but one would expect the attenuation constant to be increased due to the slight rise in level involved. At higher frequencies, the meniscus effect would seriously affect the results, owing to the narrow guide sections which would be employed. For these frequencies, it would be necessary to fill the guide with a movable polythene filling, and to make the necessary field measurements at two points in the polythene filled part. The liquid would then be inserted through suitable slots into the region below the polythene slide. The detectors would, of course, need to move with the polythene.

A further series of measurements was carried out on solutions of water or ethyl alcohol in dioxan. An attempt to measure the absorption of pure alcohol was made, but it was found impossible to reproduce results for absorption measurement, more closely than within ±5%. This was no doubt due to the fact that the experimental curve showed but one maximum, owing to the relatively low dielectric constant of ethyl alcohol at these frequencies.

Table 2. Variation of Values of Absorption and Dielectric Constant in Mixtures of Water and Dioxan at 21°c (frequency 3×10^9 c/s)

Molar fraction of H_2O in $(C_2H_4)_2O_2$	λ_L (cm)	ϵ' (±1%)	ϵ''	α'
0·0000	7·38(4)	2·26	—	—
0·0053	7·29(6)	2·31	—	—
0·0109	7·24(0)	2·34	—	—
0·0163	7·18(0)	2·37	—	—
0·0216	7·09(6)	2·42	—	—
0·0269	7·05(5)	2·44	—	—
0·0523	6·76(0)	2·62	0·04(2)	0·009
0·0765	6·510	2·79	0·06(8)	0·014
0·0995	6·270	2·97	0·09(2)	0·018
0·1214	6·060	3·15	0·13(0)	0·026

Table 2 *a*. Values obtained by Cook at 18°c for Mixtures of Water in Dioxan at $\lambda = 10$ cm

Molar fraction of H_2O in $(C_2H_4)_2O_2$	0·05	0·10	0·15
ϵ'	2·60	3·00	3·40
ϵ''	0·052	0·126	0·230

The investigation of the behaviour of very dilute solutions of water and alcohol in dioxan was undertaken in order to check the operation of the method in the small absorption region. The results obtained are tabulated in tables 2 and 3. In these tables, the molar fraction of water or alcohol in the dioxan mixture is given, together with the depth of liquid required to provide a significant value for either the dielectric constant, or the absorption coefficient. For simplicity, the second minimum of the experimental curve was used, consequently the depth of liquid in this case corresponded to one wavelength of the radiation in the liquid filled guide. Table 2 *a* gives some recent values by Cook, for comparison with table 2.

In the case of water in dioxan, repeatable results could be obtained only when the molar fraction of water was in excess of 0·05, giving α' (the experimentally determined attenuation constant of the liquid filled part of the guide) a value of 0·009 cm^{-1}.

Table 3. Variation of Values of Absorption and Dielectric Constant in Mixtures of Ethyl Alcohol and Dioxan at 21°c (frequency 3×10^9 c/s)

Molar fraction of C_2H_5OH in $(C_2H_4)_2O_2$	λ_L (cm)	ϵ ($\pm 1\%$)	ϵ'	α'
0·0000	7·38(4)	2·26	—	—
0·0263	7·17(5)	2·37	—	—
0·0438	7·03(5)	2·46	0·03(5)	0·007
0·0614	6·89(0)	2·53	0·04(1)	0·009
0·0965	6·631	2·69	0·06(7)	0·014
0·1665	6·230	3·00	0·14(7)	0·029

In the case of ethyl alcohol, the minimum value of α' which could be repeated was 0·007 cm^{-1}, requiring a molar fraction of alcohol in excess of 0·04.

These figures agree well with the theoretical lower limit of 0·005 for α', especially since the latter was derived on the assumption that a liquid depth of 10 cm could be used. In the above work, however, no depth greater than 7·40 cm was employed.

§ 6. CONCLUSION

The method developed is well suited to investigation of the electrical properties of aqueous solutions in the 10 cm wavelength region, and can be used for the measurement of absorption in liquids where absorption coefficients are in excess of 0·005 cm^{-1}. The technique is simple, and calculations are straightforward provided aqueous solutions are employed, and numerous minima are available for the purpose of evaluating α' by the direct method described. Under these conditions, an accuracy within $\pm 1\%$ is obtained.

ACKNOWLEDGMENTS

The author expresses his thanks to Professor H. T. Flint and Dr. G. Williams for suggesting the development of the method, and to members of the technical staff of the physics department of Bedford College for their help in making the apparatus.

REFERENCES

COLE, K. S., and COLE, R. H., 1941, *J. Chem. Phys.*, **9**, 341.
COLLIE, C. H., RITSON, D. M., and HASTED, J. B., 1948, *Proc. Phys. Soc.*, **60**, 145.
COOK, H. F., 1951, *Thesis*, University of London.
DEBYE, P., 1929, *Polar Molecules* (New York : The Chemical Catalog Co.).
FLINT, H. T., and PINCHERLE, L., 1943, *Proc. Phys. Soc.*, **55**, 329.
LITTLE, V. I. 1953, *Proc. Phys. Soc.* B, **66**, 185.
MONTGOMERY, C., 1947, *Technique of Microwave Measurements* (M.I.T. Radiation Laboratory Series, Vol. 2).
SAXTON, J. A., 1952, *Proc. Roy. Soc.* A, **213**, 473.
SAXTON, J. A., and LANE, J. A., 1952, *Proc. Roy. Soc.* A, **213**, 400.

An Analogue Computer employing the Principle of the Kelvin Bridge

By V. I. LITTLE
Bedford College, London

Communicated by H. T. Flint ; MS. received 22*nd September* 1952, *and in final form* 18*th November* 1952

Abstract. A computer is described which permits the rapid evaluation of α from the equation $\cosh 2\alpha S_1/\cosh 2\alpha S_2 = A$, where S_1, S_2 and A are known constants. An accuracy of within $\pm 0{\cdot}2\%$ may be achieved over the working range of the design.

---

§ 1. INTRODUCTION

In the course of recent work by the author on the measurement of the dielectric properties of liquids (Little 1953), it became necessary to devise a rapid method for evaluating α from the equation

$$\cosh 2\alpha S_1/\cosh 2\alpha S_2 = A, \qquad \ldots\ldots(1)$$

where S_1, S_2 and A were experimentally determined quantities. Although eqn. (1) may be solved arithmetically without difficulty, much labour is involved if numerous sets of experimental data require interpretation. An analogue computer was eventually made which incorporated the principle of the Kelvin bridge to achieve a null balance condition.

§ 2. THEORY OF THE COMPUTER

Consider the equation

$$y = \log \cosh x. \qquad \ldots\ldots(2)$$

If y_1 and y_2 are two values of y corresponding to the values x_1 and x_2 of x, then

$$y_1 - y_2 = \log \cosh x_1 - \log \cosh x_2. \qquad \ldots\ldots(3)$$

Let y_1 and y_2 assume a series of values such that the right-hand side of eqn. (3) is always constant. Of the corresponding series of pairs of values for x_1 and x_2, there is a unique pair whose ratio is S_1/S_2 where S_1 and S_2 are known quantities. Hence if x_1 and x_2 are equated to $2\alpha S_1$ and $2\alpha S_2$ respectively, the logarithm of eqn. (1) may be written

$$\cosh 2\alpha S_1 - \cosh 2\alpha S_2 = \log A = y_1 - y_2. \qquad \ldots\ldots(4)$$

In fig. 1 (*b*), the curve gh represents the graph of eqn. (2), y plotted vertically and $2\alpha S$ horizontally. The lines ab and cd represent the equations $y = y_1$ and $y = y_2$, the distance e_1e_2 corresponds to $\log A$ from eqn. (4), and the intervals e_1f_1 and e_2f_2 correspond to $2\alpha S_1$ and $2\alpha S_2$ respectively. For given values of S_1, S_2 and A, the unknown constant α can be found if the lines ab and cd are moved together up the axis of y, until the ratio of the lengths e_1f_1 and e_2f_2 is the same as the ratio of the known quantities S_1 and S_2. If the lines ab, cd and gh in fig. 1 (*b*) are replaced by wires possessing the same resistance per unit length, the adjustment of the ratio of the lengths e_1f_1 and e_2f_2 to any desired value may be made by finding the ratio of the resistances of these portions of ab and cd, using a Kelvin bridge. The broken lines in fig. 1 (*b*) indicate the extra circuit components required for this purpose, and the Kelvin bridge circuit of fig. 1 (*a*) is provided so that a comparison may be made.

The balance condition for the Kelvin bridge circuit of fig. 1 (*a*) is that $P/R = Q/S = X/Y$, where P, Q, R and S are large compared with X and Y. It follows, therefore, that when the resistance ratios P/R and Q/S have both been set equal to the ratio S_1/S_2, the ratio X/Y will also be equal to S_1/S_2 when the bridge is balanced. For a given set of values for S_1, S_2 and A, a value of α may be obtained from the device as follows: (i) set the separation of ab and cd to equal log A, (ii) adjust the resistances P, Q, R and S so that $P/R = Q/S = S_1/S_2$, (iii) move ab and cd together up the axis of y until the galvanometer of the Kelvin bridge circuit indicates a null, (iv) divide the value of e_1f_1 at the balance point by $2S_1$—or e_2f_2 by $2S_2$—to obtain α.

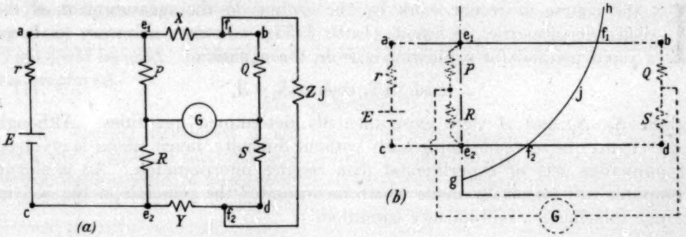


Fig. 1. (*a*) Circuit diagram of Kelvin bridge for comparison of two low resistances, X and Y; (*b*) the adaptation of the Kelvin bridge to the analogue computer.

§ 3. FACTORS AFFECTING THE ACCURACY

This form of analogue computer may be made to yield an accuracy within ± 0·2%, depending on the range of values of A, S_1 and S_2 it is desired to cover. If these quantities are grouped into narrow ranges of values, there seems to be no reason why the quoted accuracy figures should not be improved.

The inherent accuracy of the device is governed principally by the following factors: (i) The accuracy with which a conducting wire can be made to coincide with the shape of a selected portion of the curve $y = \log\cosh 2\alpha S$. This includes variations in the thickness of the wire, and location errors. (ii) The uniformity of the slide wires ab and cd. These must be straight, parallel, of constant resistance per unit length and of constant diameter. The contact resistances at the points of sliding contact, and the resistance of the portion of the curve gh intercepted between ab and cd, constitute the resistance Z of fig. 1 (*a*), and this does not appear in the balance condition. The effect of the variation of surface resistance of the wires ab and cd at various points is to vary the value of Z, and cannot influence the balance point. The resistances of those portions of the slide wire beyond f_1 and f_2 in fig. 1 (*b*) are in series with the large resistances Q and S respectively and may thus be ignored.

§ 4. Brief Constructional Details

The computer consisted of two main parts, the master curve unit in which a piece of wire was bent to the shape of the curve $y = \log \cosh 2\alpha S$ and suitably fixed in position, and the adjustable slide wire system. The master curve unit consisted of an accurate large scale plot of the equation $y = \log \cosh 2\alpha S$ for values of $2\alpha S$ running from zero to 2. A scale of 25 inches to the unit for y, and 10 inches to the unit for x was employed. This curve was fixed below a thin perspex sheet, and a suitably shaped 15 s.w.g. brass wire was cemented to the top of the sheet using Perspex cement. The wire was located accurately using templates of Bristol board which had been modelled on the master curve and corrected for the thickness of the wire. Several layers of cement were applied until the wire was encased in a Perspex ridge. After the cement had set hard, the Perspex ridge was rubbed down until the top surface of the wire was exposed. The complete unit was then mounted on a drawing board so that the y axis of the curve was parallel to two boxwood rules, which were mounted along a pair of opposite edges of the board, see fig. 2.

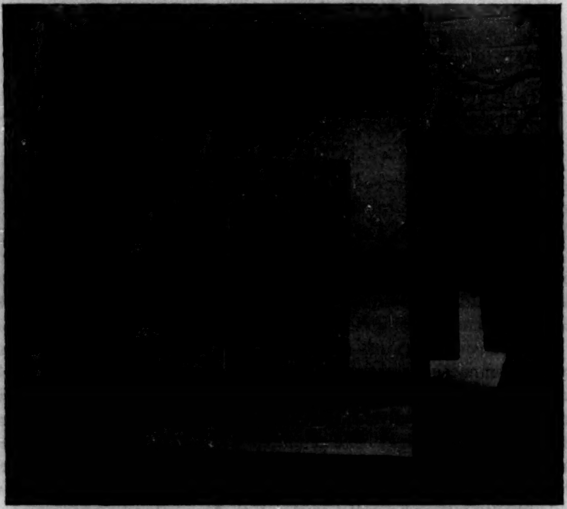


Fig. 2. Plan view of computer showing mounting of the squares and master curve.

The slide wire system consisted of two 1·34 mm diameter molybdenum wires which were mounted under the straight edges of two T-squares and located there by thin paxolin strips. The locating edges of the T-squares engaged with the boxwood rules on the drawing board of the master curve unit so that the slide wires made contact with the master curve and were kept parallel to the $2\alpha S$ axis of the latter. The separation of the slide wires at any desired distance representing $\log A$ was achieved by using a thick Perspex sheet as a spacer. Reference to fig. 2 shows this sheet bolted at one edge to one T-square and to the other square by bolts projecting through two parallel slots. The points where the slide wires crossed the y axis of the master curve were located by fine phosphor bronze spring contacts which could be adjusted to their correct setting once the T-squares were in position on the drawing board. These contacts corresponded to the points e_1 and e_2 of fig. 1 (*b*) and fig. 3 shows the details of one of them.

The choice of wires for the two parts of the computer was governed by the following considerations. In the case of the master curve, uniformity of diameter was essential in order to avoid errors in the location of the cross-over points. The piece of 15 s.w.g. hard brass wire employed did not vary in diameter by more than 1% over its length, and this degree of variation could not possibly introduce errors in location exceeding 0·1%. The slide wires had to conform to the condition that their mean resistances per unit length were the same for lengths varying from 6 inches to 20 inches. Many pairs of wires of different materials were calibrated using a Kelvin bridge, and the molybdenum wires were selected, the maximum deviations between their measured resistances per unit length amounting to only one part in 500. The diameters of the molybdenum wires were 1·34 mm each, and the average diameter of the brass wire was 1·86 mm. With wires of these sizes the conditions of uniformity were easily satisfied and their added rigidity was advantageous. The location of the cross-over points f_1 and f_2 once a balance had been obtained was in no way complicated by the thickness of the wires. The method employed was to sight a cross-over point along the edge of a set square, the locating ledge of which was engaged with an edge of the drawing board parallel to the $2\alpha S$ axis of the master curve. With a lamp arranged behind the cross-over point on the side remote from observation, the actual point of contact between the slide wire and the master curve could be seen as a small black spot against a bright background. The value of $2\alpha S$ could then be read off directly at the point where the set square traversed the $2\alpha S$ axis.

Fig. 3.

A series of checks using calculated data for A, S_1 and S_2 showed that the computer could give accurate values for α consistent to $\pm 0{\cdot}2\%$, a figure small compared with the accuracy of the experimentally determined values of A, S_1 and S_2 which the instrument was designed to interpret.

Reference

Little, V. I., 1953, *Proc. Phys. Soc.* B, **66**, 175.

The Refractive Indices of Water Vapour, Air, Oxygen, Nitrogen, Hydrogen, Deuterium and Helium

By L. ESSEN
Electricity Division, National Physical Laboratory

MS. received 10*th November* 1952

Abstract. The method described by Essen and Froome has been used to measure the refractive indices of water vapour, air, oxygen, nitrogen, hydrogen, deuterium and helium at a frequency of 9200 Mc/s. The value of $(n-1)10^6$ obtained for water vapour at 10 mm Hg pressure and 20°c is 60·7 which is the same as the value obtained at 24 000 Mc/s, showing that the absorption line at 22 000 Mc/s has a negligible effect on refractive index.

The values obtained for air, oxygen and nitrogen also agree with those previously obtained within the limits of error and therefore confirm the earlier results.

The values of $(n-1)10^6$ obtained for hydrogen and deuterium are higher than those calculated by Ishiguro *et al.* by 2% and 3% respectively. The calculated values were expected to be a few per cent lower than the true values, and they are therefore supported by our experimental results.

§ 1. Introduction

A series of measurements of the refractive indices of air, its main constituents, and water vapour has been described by Essen and Froome (1951) and further results are now given. They appeared to be required for the following reasons : (i) The earlier measurements were made at a frequency of 24 000 Mc/s (wavelength 1·25 cm) because this was the frequency at which the values were needed for work on the velocity of propagation of electromagnetic waves. There is however an absorption line of water vapour near 22 000 Mc/s (wavelength 1·35 cm) and although this is not expected to alter the refractive index appreciably it was thought that an experimental check at a frequency well removed from the line would be useful. (ii) Since the above publication several other sets of results giving appreciably different values, though admittedly with lower accuracy, have been announced. It is most important for practical applications that the refractive index of moist air, in particular, should be known with an accuracy within 1 part in 10^7, and although we saw no reason to doubt our results it was felt that additional measurements with an entirely new equipment and at a different frequency would help to establish their validity or reveal any unsuspected sources of systematic error. (iii) A recent paper by Ishiguro *et al.* (1952) shows that the polarization and dielectric constant of hydrogen and deuterium can be calculated with a high accuracy. The spread of previous experimentally determined values of $(n-1)10^6$ is about 6% and the calculated values are estimated to be accurate to within a few per cent. A more accurate experimental value will therefore be useful as a guide to the theoretical physicist. The difference between the values for hydrogen and deuterium is important as it is explained only by an advanced theory of molecular structure, and gives the value of $\alpha' = d\alpha/dr$ where α is the polarization and r the internuclear distance.

§ 2. METHOD AND APPARATUS

The method employed was the same as that described by Essen and Froome (1951), the resonant frequency of a cavity resonator being measured when it was evacuated and when filled with the gas under test. The refractive index is the ratio of these frequencies. To obtain the high sensitivity required the resonator was connected in a waveguide bridge energized by a stable source of oscillation the frequency of which could be measured with an accuracy of 1 part in 10^8.

For the present measurements the apparatus was designed to operate at 9200 Mc/s. The resonator was constructed from invar and copper plated. It resonated in the H_{019} mode with a *Q*-factor of 40 000 when matched to the waveguide. The matching was made as precise as possible by a very gradual increase of the size of the coupling hole so that the maximum sensitivity in the bridge could be obtained. It had been used previously in an investigation on stable microwave oscillators (Essen 1953) and is fully described in that work. Improved components together with the greater power available from the source in this wave-band enabled the precision of setting to resonance to be increased to 1 part in 10^8.

The arrangements for evacuating the resonator and for filling it with the gas under test were unchanged.

§ 3. RESULTS

The results are given in the table together with those obtained by Essen and Froome at 24 000 Mc/s.

Refractive Indices of Water Vapour and Dry Gases

(Dry gases at 0°c, 760 mm Hg; water vapour at 20°c, 10 mm Hg)

	$(n-1)10^6$				
Gas	Essen and Froome 24 000 Mc/s	Present values 9000 Mc/s	No. of measurements	Standard deviation	Limits of error*
Water vapour	60·7 ±0·1	60·7	20	0·15	±0·2
Air (CO_2–free)	288·15±0·1	288·10	11	0·03	±0·1
Oxygen	266·4 ±0·2	266·2	9	0·04	±0·2
Nitrogen	294·1 ±0·1	294·1	5	0·04	±0·1
Hydrogen		136·0	18	0·04	±0·2
Deuterium		134·8	5	0·2	±0·3
Helium		35·0	11	0·03	±0·2

* Including systematic errors.

The measurements on water vapour were made at pressures between 5 mm Hg and 15 mm Hg and at temperatures between 16°c and 30°c. From each observation the value of B was determined from the Debye equation:

$$\epsilon - 1 = n^2 - 1 = p\left[\frac{A}{T} + \frac{B}{T^2}\right], \qquad \ldots\ldots(1)$$

A being taken as $2{\cdot}00 \times 10^{-4}$ for reasons discussed in § 4. From the average values obtained for B the value of $n-1$ for the stated conditions of temperature and pressure was calculated.

The main experimental difficulty in the measurements on water vapour was the progressive change of pressure after the tap to the water reservoir had been closed. This was due to the adsorption of the vapour by the walls of the glass tubes and pressure container and also possibly by the metal surfaces including the inside walls of the resonator. By making a number of simultaneous readings of pressure and resonant frequency it was found that the same value was obtained for refractive index within the limits of experimental error, showing that even if water vapour was being absorbed by the walls of the resonator the results were not thereby affected to an appreciable extent. The final measurements were made in this way without waiting for steady conditions to be attained on either the evacuation or the filling of the resonator. In effect the change of refractive index for a measured change of pressure was measured, and it was considered that the state of the walls would be more nearly the same for the two conditions than it would be if the resonator were pumped out hard between the measurements. The precision of setting was slightly reduced however by this technique and this may account for the standard deviation being somewhat higher than in our previous results.

It may be mentioned in connection with this effect that Birnbaum and Chatterjee (1952) concluded that water vapour was adsorbed by the walls of the resonator but that the effect on resonant frequency was negligible because the electric field is essentially zero at the walls for the H_{01} mode.

The other gases except air and deuterium were obtained in cylinders from the British Oxygen Co. Their stated purities are as follows : Oxygen, 99·5% main impurity nitrogen ; Nitrogen, impurity less than 1×10^{-6} ; Hydrogen, 99·95% main impurity nitrogen ; Helium, 99·8% main impurity neon. The results have been corrected for the main impurities and the limits of error allow for their being twice the stated amount. The deuterium was supplied, in two samples, by the Clarendon Laboratory, Oxford; the first sample was tested there by a mass spectrometer and was found to be 99·32% pure, there being 0·68% of hydrogen present.

The measurements were made at temperatures of about 20°C and at atmospheric pressure ; and the results have been reduced to standard conditions by the use of appropriate interpolation formulae. For dry air and its main constituents equation (2) of Essen and Froome (1951) is used as this was established by the experimental work of Barrell and Sears (1939). The results for the other gases are reduced to S.T.P. by means of the ideal gas formula because the interpolation equations have not been so firmly established. Because of the uncertainty in the law the error in the values of $(n-1)10^6$ is, however, not likely to be more than 0·02% and the correct measured values can be obtained at 20°C by again using the ideal gas law.

§ 4. DISCUSSION

The results for water vapour show that the value of $(n-1)10^6$ is the same, within the experimental uncertainty of 0·2%, at 24 000 Mc/s and 9000 Mc/s, and therefore that the effect of the absorption line at 22 000 Mc/s is, as predicted, negligibly small. The actual value is strongly supported by this second series of measurements although it is nearly 2% lower than that obtained by Birnbaum and Chatterjee (1952) and also than the average given by them of a number of other results. It is pointed out in their paper that the polarization due to infra-red absorption has been measured by Greenfield and Brown (1950) and this contribution has therefore been added to that due to optical absorption to give a value for A in eqn. (1) of $2{\cdot}0\times10^{-4}$ instead of the $1{\cdot}725\times10^{-4}$ used by Essen and Froome (1951). The value of B is thereby reduced from 0·991 to 0·982 and that of the dipole moment from $1{\cdot}839\times10^{-18}$ e.s.u. to $1{\cdot}832\times10^{-18}$ e.s.u. Our interpolation formula is still valid for the stated limits because for small temperature ranges the different values of A and B compensate each other. It would however be more accurately expressed as

$$(n-1)10^6=\frac{100p}{T}\left[1+\frac{4912}{T}\right](1+2{\cdot}4\times10^{-5}p). \qquad \ldots\ldots(2)$$

The results for dry air, nitrogen and oxygen agree with our previous values within our limits of error and serve to demonstrate the accuracy and reliability of the method. The standard deviation was found to be the same as before in spite of the improved precision of setting, and this suggests that the errors are mainly associated with the determination of the temperature and pressure of the gas and possibly with its dryness.

The close agreement between the two sets of results coupled with the care taken to assess systematic errors give a strong indication that the values given are correct to within the stated limits and that the results obtained by a number of previous investigators are in error even though the discrepancies are in some cases considerably greater than the estimated limits of error. Helium was measured because the molecular structure is simple and the polarizability can therefore presumably be calculated although no values have come to our notice. The value obtained is in good agreement with that of Birnbaum, Kryder and Lyons (1951) and with various optical determinations. The value given by Cuthbertson and Cuthbertson (1932) for example is 34·6 for $(n-1)10^6$. This value was used by Gabriel (1952) to calibrate his microwave equipment which was then used to measure air, argon and neon. He gives the value for helium as 1·000 034 613 but it should be remembered that the best optical measurements in spite of the large number of figures sometimes quoted are uncertain in the 1×10^{-7} figure and many are uncertain in the previous figure. The value for argon found by Essen and Froome (1951) was 4×10^{-7} higher than the Cuthbertsons' value and that for helium reported in the present paper is higher than theirs by the same amount. A correction of this amount applied to Gabriel's values for air and argon would bring them to close agreement with ours.

The main interest in the results for hydrogen and deuterium rests in their comparison with theoretical values. Bell (1942) shows that values of the polarizability α of two isotopic molecules are not necessarily the same and he derives the difference $(\alpha_{\mathrm{H}}-\alpha_{\mathrm{D}})/\alpha_{\mathrm{H}}$ from experimental work in the optical region on molecules containing H_2 and D_2. Ishiguro *et al.* (1952) have summarized the calculated values of the polarizability of hydrogen. They vary by 25% but were based on rather simple wave functions of the molecule and for one or several fixed values of intermolecular distance. Ishiguro *et al.* have extended the calculations using the accurate eleven-terms wave function obtained by James and Coolidge (1933). They obtained the values

$(n-1)10^6 = 133{\cdot}3$ for hydrogen and 130·8 for deuterium and point out that as the values are still increasing with the number of terms taken they will be rather lower than the true values. It will be seen from the table that our measured values are 2% and 3% higher respectively and that the percentage difference between them is 1% compared with the 1·86% of Ishiguro *et al.* and the 1·31% obtained by Bell.

ACKNOWLEDGMENTS

The work described above has been carried out as part of the research programme of the National Physical Laboratory and this paper is published by permission of the Director of the Laboratory.

REFERENCES

BARRELL, H., and SEARS, J. E., 1939, *Phil. Trans. Roy. Soc.* A, **238,** 1.
BELL, R. P., 1942, *Trans. Faraday Soc.*, **38,** 422.
BIRNBAUM, G., and CHATTERJEE, S. K., 1952, *J. Appl. Phys.*, **23,** 220.
BIRNBAUM, G., KRYDER, S. J., and LYONS, H., 1951, *J. Appl. Phys.*, **22,** 95.
CUTHBERTSON, C., and CUTHBERTSON, M., 1932, *Proc. Roy. Soc.* A, **135,** 40.
ESSEN, L., 1953, *Proc. Instn. Elect. Engrs.*, Pt. III, **100,** 19.
ESSEN, L., and FROOME, K. D., 1951, *Proc. Phys. Soc.* B, **64,** 862.
GABRIEL, W. F., 1952, *Proc. Inst. Radio Engrs.*, N.Y., **40,** 940.
GREENFIELD, M. A., and BROWN, F. W., 1950, *J. Opt. Soc. Amer.*, **40,** 643.
ISHIGURO, E., ARAI, T., MIZUSHIMA, M., and KOTANI, M., 1952, *Proc. Phys. Soc.* A, **65,** 178.
JAMES, H. M., and COOLIDGE, A. S., 1933, *J. Chem. Phys.*, **1,** 825.

Absorption Spectra of Lead Sulphide at Different Temperatures

By W. PAUL and R. V. JONES
Department of Natural Philosophy, University of Aberdeen

MS. received 11*th August* 1952

Abstract. The absorption spectrum of lead sulphide, as measured between 1 and 10 microns, shows a sharp decrease (in some specimens from 100 cm^{-1} to 5 cm^{-1}) in the absorption coefficient at a wavelength corresponding to the limit of photoconductivity. In many galena crystals additional absorption exists that masks completely the sharp absorption edge. The absorption edge and the photoconductive limit change their position similarly with temperature, the rate of shift of the edge being $+3{\cdot}7 \times 10^{-4}$ ev/° K between 90° K and 400° K. The shift is in the opposite direction to that observed in the elementary semiconductors silicon and germanium. At long wavelengths the absorption coefficient increases with the square of the wavelength, but there is a large discrepancy between the magnitude of the measured absorption coefficient and that calculated on the free electron theory. The long wavelength absorption increases with the temperature; this may correspond either to an increase in the number of carriers or to a decrease in their mobility.

§ 1. INTRODUCTION

THE long wavelength limit of photoconductivity in films of lead sulphide lies near 3 microns wavelength at room temperature and moves to longer wavelengths as the temperature is lowered.

At this limit a marked decrease in absorption coefficient might be expected, followed by relative transparency at longer wavelengths; such a decrease occurs, for example, in the cases of silicon and germanium. Prior to the present investigation, however, no such decrease had been observed, either for single crystals of galena or for thin films of lead sulphide (Gibson 1950) although in the latter case the absorption coefficient fell from 10^5 cm^{-1} to 10^4 cm^{-1} between 0·4 and 1·5 micron wavelength. It has therefore been desirable, for a better understanding of the photoconductive processes in lead sulphide and similar materials, to establish whether the previous failures to discover a change in transparency at the photoconductive limit were due to an intrinsic property of the pure material, or to a masking absorption by some accidental impurity.

Recently, we have had the good fortune to examine some crystals of natural galena left in this Department by Sir George Thomson after his experiments here on electron diffraction between 1927 and 1930. These crystals, which are of unknown origin and which possess no impurity detectable by x-ray analysis, show a sharp change to a remarkably low value of absorption coefficient (5 cm^{-1}) at the photoconductive limit. This fact has already been summarily reported (Paul, Jones and Jones 1951). The position of the absorption edge moves to shorter wavelengths as the temperature of the crystal is increased.

By courtesy of the gentlemen mentioned later, we have since examined natural galena from many sources and find that the transparency beyond 3 microns is a rare property. Of all the crystals tried, only those from Wisconsin have been as good as those which we originally examined. Crystals from the Harz Mountains, Cornwall and Lanarkshire also show an absorption edge but have a higher absorption coefficient (order of 25 cm^{-1}) (see table 1). All the other crystals examined have absorption coefficients higher than can be measured by our spectrometer.

Table 1

Specimens that have *not* shown transmission.

Foxdale, Isle of Man.
Alston, Cumberland.
Matlock, Derbyshire.
Rhosemoor Mine, Flintshire.
Ivigtut, Greenland.
Caldbeck Fells, Cumberland.
Keswick, Cumberland.
Wanlockhead, Dumfriesshire.
Lunehead Mines, N.W. Yorks.
Mogador, Morocco.

Specimens that *have* shown transmission.

Cornwall—Low.
Harz—Low.
Wisconsin—Transparency comparable with specimen reported.
Leadhills, Lanarkshire—Low.

Since our previous communication (Paul, Jones and Jones 1951) measurements on the position and shift of the absorption edge in single crystals of lead sulphide have been reported by Clark and Cashman (1952) and by Gibson (1952).

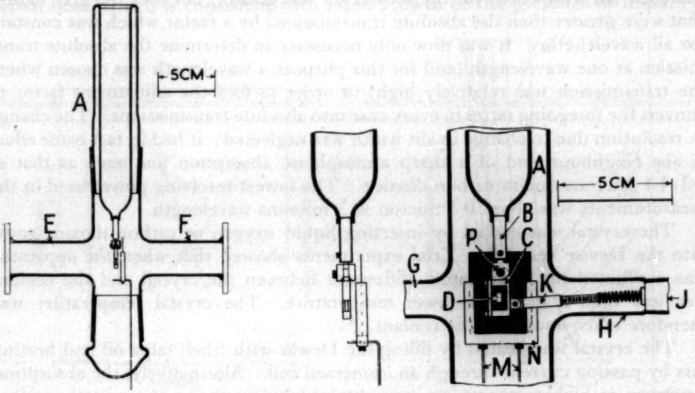


Fig. 1. Apparatus used in absorption coefficient measurements at different temperatures. Explanation of symbols in text.

§2. EXPERIMENTAL METHODS

The lead sulphide crystal was kept at a fixed temperature in the apparatus shown in fig. 1. A Dewar vessel A had, sealed through the bottom of its inner wall, a copper rod B carrying a holder C for the crystal D. The crystal was surrounded by copper strips and any unavoidable spaces between crystal and walls were filled by copper foil so that any radiation from the Nernst filament reaching the spectrometer slit must have passed through the crystal.

The apparatus had four sidearms and a removable base. Two opposite arms E and F were sealed with windows of lithium fluoride which could be removed for measurements at room temperature beyond 6 microns wavelength. The arm G was used for evacuating the apparatus. The male part of the joint J to the arm H had attached to its end a spring and a screw that located in a tapped brass rod K hinged to the crystal holder C. The base of the apparatus carried two upright thick tungsten wires M supporting a holder and screen N. This screen was situated just in front of the crystal and had a small rectangular or circular aperture restricting the incident beam so that it had to pass through the crystal, which was seldom larger than 0·8 cm by 0·5 cm. Rotation of the joint J moved the crystal holder perpendicularly out of the beam by turning it about the pin P.

The apparatus was mounted so that the crystal was at the pre-slit focus of a Hilger D.209 infra-red spectrometer. The emission spectrum of the Nernst filament was determined over the whole range of wavelengths concerned; the sensitivity of the thermopile and subsequent amplifier was kept constant. Then the energy transmitted through the crystal was measured over the same wavelength range, at the same thermopile and amplifier sensitivity but with a slit width greater than in the previous case and wide enough to give accurately relative measures of the energy transmitted at different wavelengths. The ratios of the two sets of measurements did not give the absolute transmission of the crystal, since the slit width used for the second set was greater, but they did give figures that were greater than the absolute transmissions by a factor which was constant for all wavelengths. It was now only necessary to determine the absolute transmission at one wavelength (and for this purpose a wavelength was chosen where the transmission was relatively high) in order to find the adjustment factor to convert the foregoing ratios in every case into absolute transmissions. The change in resolution due to change in slit width was neglected; it had in fact some effect in the neighbourhood of a sharp atmospheric absorption line such as that at 4·3–4·4 microns due to carbon dioxide. The lowest resolving power used in the measurements was about 0·1 micron at 3 microns wavelength.

The crystal was cooled by inserting liquid oxygen or carbon dioxide snow into the Dewar vessel A. Trial experiments showed that when the apparatus was evacuated the temperature difference between the crystal and the coolant was less than 2° c at the lower temperature. The crystal temperature was therefore taken as that of the coolant.

The crystal was heated by filling the Dewar with Shell talpa oil and heating this by passing current through an immersed coil. Alternatively, the absorption spectrum at high temperatures was obtained by mounting the crystal inside a horizontal furnace having lithium fluoride windows. The furnace was long enough for the crystal to 'see' hot furnace wall over a large solid angle. The shielding and measuring arrangements were similar to those in the apparatus described above.

§ 3. RESULTS

(i) *Room Temperature*

Under the conditions of the experiments it is possible to neglect internal reflections in the crystal, so that the transmitted intensity I of radiation incident with intensity I_0 is given by $\ln(I_0/I)+2\ln(1-r)=\mu t$ where t is the thickness,

μ the absorption coefficient, and r the reflectivity, of the crystal. This relation was used to determine the values of μ and r from measurements on eight thicknesses of crystal from 0·05 to 0·28 cm in the wavelength range of 3 to 10 microns. The results are shown in table 2.

Table 2. Values of Absorption and Reflection Coefficients at Room Temperature

λ (μ)	3·25	3·5	4·0	5·0	6·0	7·5	8·5	10·0
μ (average)	12·1	5·85	5·25	5·9	7·0	10·3	12·5	18·2
r (average)	0·424	0·443	0·414	0·402	0·398	0·373	0·390	0·369

Average value for $r = 0{\cdot}40$, hence refractive index $n = 4{\cdot}4$.

The value of μ is plotted against wavelength in fig. 2, which also shows a typical transmission. The absolute values of μ may be correct only to within 10%. The experimental curve shows that single crystals of lead sulphide are relatively opaque up to a wavelength near 3 microns at room temperature, that the absorption then falls sharply (in the crystals used, to an absorption coefficient of the order of 5 cm^{-1}) and rises slowly again on the long wavelength side. In other, more opaque, crystals the transmission graphs have roughly the same shape and the absorption edge occurs at the same wavelength. There is no apparent change in the transmission properties of a crystal on polishing; occasionally crystals which are opaque have a more obvious sub-crystal structure than transparent ones, which is consistent with rapid growth or the presence of impurity.

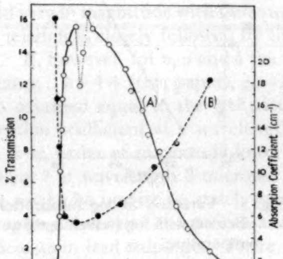


Fig. 2. Variation of the absorption coefficient of lead sulphide crystals with wavelength, and a typical transmission curve.
(A) % transmission.
(B) absorption coefficient.

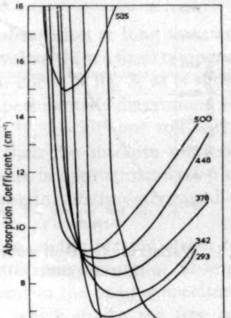


Fig. 3. Variation of absorption coefficient of lead sulphide with wavelength, for several temperatures.

The absorption might be due directly to the impurity or to some process at the grain boundaries. Attempts to determine the shape of the absorption curve at wavelengths shorter than the edge using thin crystals have so far yielded no significant results. Annealing of 'opaque' lead sulphide at a temperature near its melting point for a period of two weeks has not changed the opacity.

Figure 5 is obtained from fig. 2 by plotting the logarithm of the absorption against the logarithm of the wavelength. At the longer wavelengths the graph approximates closely to a straight line of slope 2·08, but departs from this line as the cut-off wavelength (see below) is approached.

(ii) *Other Temperatures*

Figure 3 shows the variation of absorption coefficient with wavelength for several temperatures between 90° and 585° K.

The wavelength corresponding to the sharp edge in the transmission curves can be estimated in several ways leading to slightly different results. We choose the wavelength as being that at which the transmission (as shown, for example, in fig. 2) begins to rise sharply from its low value at short wavelengths, as this point can be estimated fairly accurately. Figure 4 shows the variation in cut-off wave number with temperature; the graph is a straight line between 90° K and 400° K, after which its slope decreases considerably. Between 90° K and 400° K the shift in the edge is $+3{\cdot}7\times10^{-4}$ ev/° K. The error in this estimation is probably less than 5%; the result agrees closely with that of Clark and Cashman (1952). It is to be noted that the absorption coefficient at long wavelengths increases with the temperature.

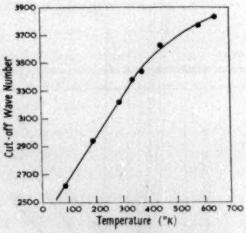


Fig. 4. Variation of cut-off wave number with temperature.

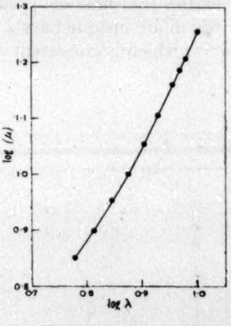


Fig. 5. Variation of log (absorption coefficient) with log (wavelength) at room temperature.

§ 4. DISCUSSION OF RESULTS

(i) *Position of Absorption Edge at Room Temperatures*

The coincidence of the absorption edge in single crystals with the photoconductive limit in lead sulphide cells would suggest that the absorption on the short wavelength side of the edge results in photoconduction. This supposition is supported by the direction and magnitude of the change in position of the edge as the temperature is changed.

The absorption edge occurs at the same wavelength in all specimens transparent enough to be measured. The absorption, and the photoconduction, may be due to electron transitions between the allowed energy levels in the pure crystal or may involve levels caused by lattice defects or impurity inclusions. If the absorption were due to electronic transitions between the top filled and first empty energy bands in the crystal the energy gap at room temperature would be approximately 0·4 ev. This value contrasts with that given by Putley and Arthur (1951) from Hall constant measurements (1·2 ev) which agrees roughly with the gap width estimated from the absorption spectra of lead sulphide films by Gibson (1950). However, these spectra showed no sharp change in absorption coefficient around 3 microns, and it is difficult to say how much of the absorption curve shape was due to the lead sulphide and how much to the structure of the film.

Although an impurity band is a possible explanation of the absorption from 1 to 3 microns, Hall constant data (Putley and Arthur 1951) give no indication of its presence.

(ii) *Magnitude of Absorption at Long Wavelengths*

The increase in absorption at long wavelengths is comparable with that observed in silicon and germanium by Becker and Fan, who discuss the absorption caused by the free carriers in the crystal. If this is so (and if a similar argument can be applied to lead sulphide) a classical treatment of the absorption process gives for the coefficient of absorption

$$\mu = \frac{4\pi}{n\mathbf{c}}\sigma_0\left(\frac{\gamma}{\nu}\right)^2 \qquad \ldots\ldots(1)$$

where σ_0 is the d.c. conductivity, n the refractive index, $\mathbf{c}$ the velocity of light, ν the frequency of the radiation and γ given by the formula $\gamma = \mathbf{e}/2\pi m^*b$, where m^* is the effective mass of an electron or hole, $\mathbf{e}$ is the electronic charge, and b the mobility.

According to this equation the coefficient of absorption at long wavelengths should vary in magnitude with the square of the wavelength, at a fixed temperature. This relation is closely followed by the absorption curve of fig. 3, as is shown in fig. 5. If, however, for n, σ and b the following experimentally determined values are taken: $n = 4{\cdot}4$ (this paper), $\sigma_0 = 36$ ohm^{-1} cm^{-1}†, $b = 330$ cm^2 volt/sec†, and m^* is assumed equal to the free electron mass, then the absolute value of the absorption coefficient at a wavelength of 8 microns is approximately 1·6 cm^{-1}. This is an order of magnitude lower than the experimentally determined value of 24 cm^{-1} at wavelength 8 microns.

It might be unwise to attach much importance, at the present stage, to the discrepancy between theoretical and experimental magnitudes of absorption coefficients in lead sulphide. There may be present in the opaque specimens of natural galena an additional absorption process which marks the free carrier absorption. These specimens are much more numerous than the transparent ones, which show much variation in transparency from one specimen to another. It would therefore be dangerous to assume that even the clearest specimen so far discovered is completely free from this possible additional absorption process. If such a process exists it will have to satisfy the foregoing condition that the resultant absorption coefficient is proportional to the square of the wavelength. The absolute experimental value of the absorption coefficient is similarly far out of agreement with theory in the cases of silicon and germanium. It is possible that the explanation of the discrepancy will be most easily afforded by these substances which have been much more fully investigated than lead sulphide.

---

† Measurements kindly made by Dr. E. H. Putley of Telecommunications Research Establishment.

(iii) *Variation of Absorption Edge and Absorption Coefficient with Temperature*

The shift in the absorption edge as the temperature is varied is in the same direction and of the same order of magnitude as the shift in the photoconductive limit of lead sulphide as determined by Moss (1949). It is similar to the shift with temperature of the edge in silicon and germanium but is in the opposite direction. The decrease in the slope of the curve of fig. 4 at very high temperatures is not, as far as is known, paralleled in silicon and germanium.

The absolute value of the absorption coefficient at long wavelengths increases as the temperature is raised. Equation (1) can be transformed to $\mu = N\mathbf{e}^3/\pi n\mathbf{c}bm^{*2}\nu^2$ where N is the carrier density. The increase in μ could then be due either to an increase in N or a decrease in b. Unfortunately, Hall constant data at elevated temperatures were not available for the specimens used in the absorption measurements, so that no definite check of theory with experiment could be carried out.

ACKNOWLEDGMENTS

We are indebted to Dr. A. E. M. Geddes, who preserved the crystals used by Sir George Thomson, and to Dr. D. A. Jones, who helped us considerably in the early stages of the investigation. We are grateful to Professor T. C. Phemister, of the Geology Department in this University, to Mr. R. C. Spiller of the Mineralogy Department at Oxford, and to the Keeper of the Mineralogical Collection at the Natural History Museum, South Kensington, who provided specimens for examination. We have also been helped by the Telecommunication Research Establishment, who measured the Hall constants reported above. The spectrometry was done on a Hilger D.209 instrument provided by a grant from the Department of Scientific and Industrial Research. The glass apparatus was made by Mr. R. Sherwood.

REFERENCES

BECKER, M., and FAN, H. Y., 1951, Semi-Conducting Materials, ed. H. K. Henisch (London : Butterworths Scientific Publications).

CLARK, M. A., and CASHMAN, R. J., 1952, *Phys. Rev.* **85,** 1043.

GIBSON, A. F., 1950, *Proc. Phys. Soc.* B, **63,** 756 ; 1952, *Ibid.*, **65,** 378.

MOSS, T. S., 1949, *Proc. Phys. Soc.* B, **62,** 741.

PAUL, W., JONES, D. A., and JONES, R. V., 1951, *Proc. Phys. Soc.* B, **64,** 528.

PUTLEY, E. H. and ARTHUR, J. B., 1951, *Proc. Phys. Soc.* B, **64,** 616.

Development of a New Method for the Absolute Determination of β-Ray Energies*

By H. CRAIG and C. F. DIETRICH
Birkbeck College, London

MS. received 26th March 1952, and in final form 1st September 1952

Abstract. Part I. The paper describes the use of a magnetic electron lens working at unit magnification, in which the image rotating property is used to produce two line images by reversal of the current through the lens windings. It is shown how the specific momentum ($H\rho$) of the focused electrons may be found from a knowledge of the angle between the two lines, the mean current and the number of turns on the lens. This is followed by a theoretical investigation of the optimum conditions necessary to obtain the highest accuracy.

A pilot experiment in which the formula for ($H\rho$) is tested, gave an accuracy within 1 in 700 for the Th F line. Further experimental work was carried out on a second lens to test the optimum conditions referred to above. It was shown that the angle between the lines could be determined to an accuracy within 1 in 30 000. It is further explained how by the use of a new optical device, and on the basis of 25 readings, the angle between the lines could be established to an accuracy within 1 in 10^5.

Part II. The second lens of which the number of turns is not known was used to investigate the specific momentum of the I, F and A lines of Th(B + C). In this application of the method the lens was calibrated by means of the I and F lines of ThB, giving the ratio of the two specific momenta, whilst the difference was obtained from x-ray data. An absolute determination of the A line is described.

PART I

§ 1. INTRODUCTION

All absolute determinations of β-ray energies have hitherto employed the so-called 'semicircular focusing method' which necessitates the production and measurement of a homogeneous magnetic field and the measurement of the radius of curvature of the electron trajectories forming the focused line. The production of a homogeneous magnetic field is difficult, and any inhomogeneity has to be allowed for by means of the Hartree correction, which entails the exact measurement of the field at many points along the electron beam. The measurement of ρ, the radius of curvature of the trajectories of the monokinetic electrons forming a line image, is also a matter of some difficulty, because of the asymmetry of electron density across the line, and because of scattering of electrons in the photographic emulsion of the recording plate. The former effect which is an aberrational one can be allowed for, but the latter effect produces a general shift of the whole line by an amount which is very difficult to estimate.†

---

* This paper contains material from theses by Craig (1952) and Dietrich (1951) which have been accepted by the University of London for the degree of Doctor of Philosophy.

For absolute determinations, the reference lines usually chosen are either the F, G and H lines of RaB or the F and I lines of ThB. The principal determinations made since 1932, of one or both of these groups of lines are listed in table 1.

It should be noted that the measurements shown in this table as 'derived' depend upon the K–L x-ray absorption limits of bismuth, and are thus not strictly absolute. The experimental work and results are described in Part II of this paper.

Table 1. Values of $(H\rho)$ in gauss cm for Main β-ray Lines of ThB and RaB

		ThB		RaB		
Worker	Year	F line	I line	F line	G line	H line
Ellis	1932	1385·8 ± 3·0	1751·0 ± 3·0	1400·4 ± 3·0	1665·9 ± 3·0	1925·5 ± 4·0
Scott	1934					1931·8 ± 0·2
Rogers	1936			1406·0 ± 0·7	1671·1 ± 0·8	1931·5 ± 1·0
Surugue	1937	1388 ± 1·4	1752 ± 2			
Arnoult	1939	1381 ± 1·4	1747 ± 2			
Siegbahn	1941					
(Derived)		1383·8 ± 4·5	1749·6 ± 5·8			
Dietrich	1951	1385·9 ± 2·0	1757·1 ± 4·0			
Lindström	1951					
(Absolute)		1388·55 ± 0·20	1754·01 ± 0·25			
(Derived)		1388·52 ± 0·06*				
Craig	1952					
(Derived)		1388·55 ± 0·30	1753·94 ± 0·40			

* From his paper (*Ark. Fys.* 1951) it is not clear what Lindström meant by assessing what he calls his 'probable experimental error' as ±0·06. If there were no error in the ratios (R) of the fields for the F and I lines, the uncertainty in the fundamental constants would alone produce a probable error of ±0·06.

§2. NEW METHOD

The new method described here was suggested to us by R. E. Siday in order to dispense with the precise measurement of the magnetic field H and the radius of curvature ρ required by the semicircular focusing method, and thus to produce a method which should be capable of greater absolute accuracy. The new method employs a symmetrical iron-free magnetic lens, in which a line source of monomolecular thickness serves as an object which is imaged at unit magnification on a photographic plate. A second image also at unit magnification but rotated in the opposite sense is produced by reversing the lens current. In general the focusing currents producing the two images will be slightly different due to the presence of the residual component of the earth's magnetic field parallel to the lens axis, the other component having been annulled by a degaussing frame.

It may be shown that the derivative of the angular coordinate of a non-skew paraxial electron with respect to the z or axial coordinate is

$$\theta'(p)=H/2(H\rho)\text{‡} \qquad \ldots\ldots(1)$$

---

† Lindstrom (1951) using a semicircular focuser uses counters to locate the line and thus avoids the scattering difficulty.

‡ This equation is given in any textbook on electron optics (e.g. Maloff and Epstein 1938).

where $\theta'(p) = d\theta(p)/dz$ (the bracketed p implying that θ is the paraxial rotation), H is the axial component of the magnetic field at a point z on the axis, and $(H\rho)$ the specific momentum. This form has been chosen since $(H\rho)$ is relativistically invariant, and the equation is thus true for all electron velocities. If (1) is now integrated between z_1 and z_2 along the lens axis

$$\theta_1(p) = [1/2(H\rho)] \int_{z_1}^{z_2} H dz + a/2(H\rho) \qquad \ldots\ldots(2)$$

where a is the axial integral of any permanent field between z_1 and z_2. Now for any solenoid

$$\int_{-\infty}^{+\infty} H dz = 4\pi n i_1 \qquad \ldots\ldots(3)$$

where i_1 is the current through the windings in e.m.u. and n is the total number of turns on the solenoid. We may thus put

$$\int_{-\infty}^{+\infty} H dz = 4\pi n i_1 = \int_{z_1}^{z_2} H dz + 4\pi n i_1 K_H \qquad \ldots\ldots(4)$$

where K_H is a constant of the solenoid for given z_1 and z_2. Thus substitution from (4) in (2) gives

$$\theta_1(p, z_1, z_2) = 2\pi n i_1(1 - K_H)/(H\rho) + a/2(H\rho) \qquad \ldots\ldots(5)$$

where $\theta(p, z_1, z_2)$ is the paraxial image rotation between the axial points z_1 and z_2. If the current is reversed and adjusted until the image is again in focus, then

$$\theta_2(p, z_1, z_2) = 2\pi n i_2(1 - K_H)/(H\rho) - a/2(H\rho). \qquad \ldots\ldots(6)$$

Therefore adding (5) and (6) and writing $i_1 + i_2 = 2i$ and $\theta_1(p,z_1,z_2) + \theta_2(p,z_1,z_2) = 2\theta$ and transposing we get finally

$$(H\rho) = 2\pi n i(1 - K_H)/\theta \qquad \ldots\ldots(7)$$

which is the required equation for $(H\rho)$.

2θ is the measured angle between the two lines whilst i_1 and i_2 are measured separately. The constant K_H can be made very small by a suitable choice of z_1 and z_2, and its value can be estimated with sufficient accuracy to make the error in the term $1 - K_H$ negligible.

Since K_H is fixed for given object and image planes, i.e. $z = z_1$, $z = z_2$, then for given $(H\rho)$

$$\delta\theta/\delta i = 2\pi n(1 - K_H)/(H\rho) \qquad \ldots\ldots(8)$$

and

$$(H\rho) = 2\pi n(1 - K_H)(i \pm \delta i)/(\theta \pm \delta\theta). \qquad \ldots\ldots(9)$$

This means that *the exact focusing current need not be found* provided the imaged lines are sharp enough for θ to be measurable accurately. The corresponding value of i/θ will then give the correct value of $(H\rho)$. This necessitates the lens having a depth of focus, and i/θ remaining constant over this focusing range.

If $i/(H\rho)$ is kept constant it follows from the general equations of motion of an electron in a magnetic field that all trajectories remain unaltered, from which it follows that for any pair of conjugate points θ is a constant. For any given $(H\rho)$ the variation of θ with i is given by eqn. (8).

§ 3. PRIMARY REQUIREMENTS FOR THE FULL UTILIZATION OF THE METHOD

Since i and n can be measured to a very high order of precision, the accuracy of the method depends essentially on the measurement of θ. It is also essential that the value of θ obtained shall be that corresponding to paraxial conditions, i.e. to an aberration-free lens. In order that the highest accuracy may be attained, the necessary conditions for satisfying the two basic requirements stated above will now be discussed. They fall into two classes, those that must be satisfied for a source emitting monokinetic electrons, and the further conditions arising when the line is accompanied by a continuous background. Lastly, the conditions governing the constancy of i/θ for a range of currents are considered, since a depth of focus is extremely valuable in obviating the necessity of finding the exact focusing current.

(i) *Conditions necessary for the Accurate Determination of θ for Monokinetic Sources*

(*a*) *Straightness of the line image.* The straightness of the image of a straight wire source depends on the distortion of the lens which may be divided into two parts, viz. isotropic distortion and anisotropic distortion.

Isotropic distortion will only curve the image if the line source does not pass through the lens axis. The presence of any small stray permanent magnetic fields, however, makes the exact location of the lens axis difficult, and so it is desirable to have no isotropic distortion.

Anisotropic distortion has the effect of over or under rotating any image point, and transforms a straight line into a figure like an integral sign. This type of distortion like the isotropic form is proportional to the cube of the distance of the image point from the axis, and being a rotational aberration is measured perpendicular to the radius sector joining the undistorted image point to the lens axis. When the lens current is reversed the sense of the distortion is reversed, and it is obvious that the presence of this aberration would render the measurement of θ inaccurate.

It has been shown (Dietrich 1951) that both of these aberrations are amenable to correction: if a symmetrical lens is used at unit magnification, then the isotropic distortion vanishes, whilst the anisotropic distortion is a minimum. Calculation showed the residual anisotropic distortion in the lens used in the pilot experiment described below to be only 1 part in 7000 of the total rotation for an image point 1 cm off the axis.

(*b*) *Condition necessary for sharp definition at edge of line image.* In order to obtain sharp definition at the edge of a line image it is very desirable that the spherical aberration should be a minimum for the lens used and, in general, to obtain the sharpest edge to the line, the lens has to be stopped down to an optimum value (see § 5 (i)).

Since the spherical aberration decreases as the solenoid windings increase in axial extent (Siday 1942) the latter were chosen to exceed the focal length. The spherical aberration for unit magnification was calculated and the intensity distribution for a point source found in the focal region, i.e. at the marginal, the disc of least confusion and the paraxial foci. By numerical integration the calculations were extended to find the intensity distribution across the image of an infinitely narrow line source and a source of finite width at the three focal positions previously mentioned.

Figures 1 and 2 show the results. It will be seen that the distributions for the thin source are generally similar to those for the source of finite width. The distributions for the latter show that the best image is to be found at the paraxial focus, where the line image is narrowest and the fall in intensity on either side of the maximum greatest. The paraxial focus is not difficult to locate since beyond it the image disappears very rapidly, whilst the disc of least confusion focus is easily identified by its greater intensity and greater width as compared with the paraxial focus.

A much sharper line is produced if the lens is stopped down until the aberration skirt surrounding the image at the paraxial focus just disappears. The aberration skirt appears to be removed, visually, when the diameter of the circle formed by the intersection of the marginal zone of rays from an axial object point with the paraxial image plane is equal to the resolution of the photographic plates used. At full aperture this diameter was 1·4 mm and at quarter aperture it was 0·02 mm, which was the approximate value of the plate resolution.

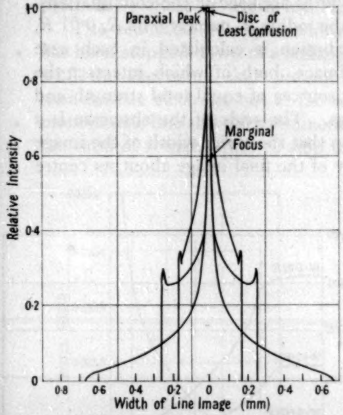


Fig. 1. Electron intensity curves for an infinitely narrow source in the focal region of a lens suffering from spherical aberration.

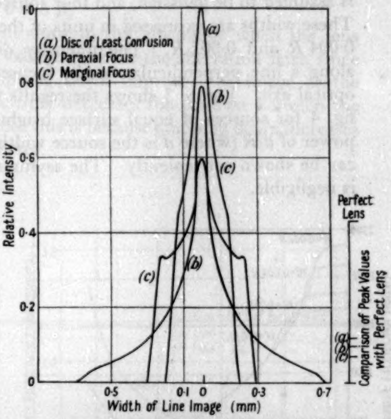


Fig. 2. Electron intensity curves for a wire 0·06 mm diameter in the focal region of a lens suffering from spherical aberration.

In the lens used in experiments to test the method the aberration skirt was removed at a numerical aperture of about 0·013 with image and object conjugate distances of about 50 cm. When used without a stop, the numerical aperture of the lenses was about 0·05. Calculation of the maximum intensity of the paraxial image showed that when the lens was stopped down from 0·05 to 0·013, the maximum image density was reduced in the ratio of 3 to 10, whereas the inverse square law gives a reduction of about 1 to 16. Stopping down still further brings a small but steady improvement of the image quality, but the image density now falls approximately according to the inverse square law, and so stopping down beyond the point where the aberration skirt is removed is of doubtful value owing to the long exposure times involved. Also, using stop values smaller than this critical value brings about, as will be seen in the next section, an undesirable diminution of contrast between the focused line and the focused continuous background.

(ii) *The Contrast of a β-Ray Line with its Accompanying Background*

In general the β-ray spectrum of a radioactive source consists of lines of discrete energy which are superposed on a continuous background. When a line source is used in a magnetic lens, the continuous background is focused continuously down the length of the lens axis, and a photographic plate placed anywhere in this focal region records an image like that of an out-of-focus line. When a line of discrete energy is focused, it is thus superposed on this 'background line', and therefore it is necessary to discuss the conditions governing the contrast between the discrete line and the background line.

Firstly, the intensity distribution due to the continuous background will be considered, the range of $(H\rho)$ covered being from 80% to 120% of the value in focus, whilst the source is assumed to be a uniform polykinetic one. The lens is assumed to be gaussian, and four source widths are used in the computations. These widths are expressed in units of the tube radius R, namely $0{\cdot}02\,R$, $0{\cdot}01\,R$, $0{\cdot}004\,R$ and $0{\cdot}002\,R$. The intensity distribution is calculated in each case along a line perpendicular to the focused image, both of which intersect the optical axis. Figure 3 shows the results for sources of equal total strength and fig. 4 for sources of equal surface brightness. The scale of the abscissae is a power of d/R (where d is the source width) so that the entire width of the image can be shown conveniently. The asymmetry of the final image about its centre is negligible.

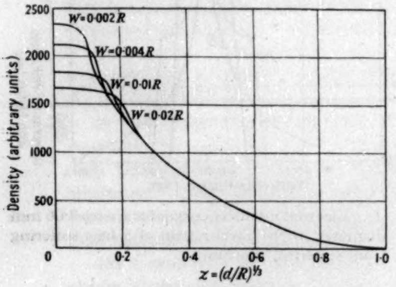


Fig. 3. Densities across the image centres for sources of equal strengths.

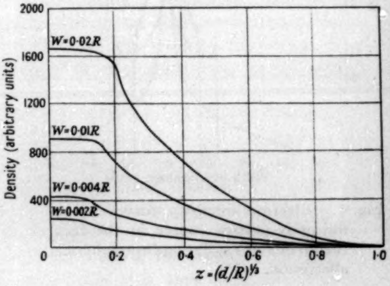


Fig. 4. Densities across the image centres for sources of equal brightness.

Two very important facts emerge from the results, the first is that the image of a uniform polykinetic line source resembles a defocused monokinetic line image; it may conveniently be called the 'background line' since it will accompany every focused image formed by conversion electrons. The second important fact is that the major contribution to the intensity of the centre of the background line comes from such a narrow range of the continuous spectrum that the shape of the continuous spectrum over this range is of no consequence. As an illustration of this it is found that 28% of the central intensity of the image produced by the continuous background is due to that part of the radiated continuous spectrum with $(H\rho)$ between 99·9% and 100·1% of the focused $(H\rho)$ and 59% is due to that part of the spectrum with $(H\rho)$ between 99% and 101%

of the focused ($H\rho$). It follows from the above that the intensity of the centre of the background line is approximately proportional to the height of the continuous spectrum curve at the ($H\rho$) being focused.

In order to compare the effects of different widths of source, the widths of the peaks at half maximum value were determined and divided by the widths of the corresponding sources. These ratios were then plotted against source widths in fig. 5. Now any conversion electrons, giving monokinetic radiation of a particular ($H\rho$) superimposed on the continuous radiation would, if focused, give a superimposed image at the image plane of width equal to that of the source. As may be seen from the figure, the ratio of the width of the monokinetic image to the width at half maximum intensity of the polykinetic line diminishes with source width, and so there is advantage in employing narrower sources, particularly those of width less than 0·006 R.

The reduction of object width will show to full advantage when the lens has been stopped down to a sufficient value to remove the aberration skirt, since with the aberration skirt present stopping down improves the image of the monokinetic electrons but at the same time decreases the contrast between the monokinetic line and the polykinetic line; this is because stopping down increases the ratio of source width to stop diameter.

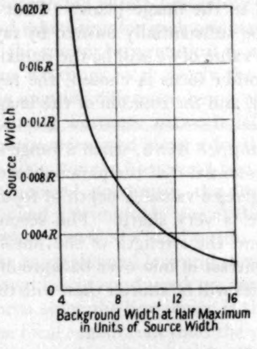


Fig. 5. Graph of image half-width for different widths of source.

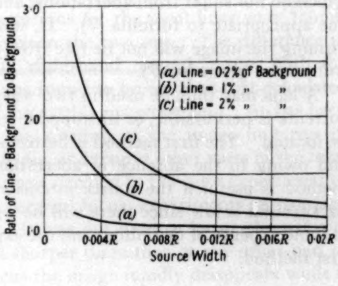


Fig. 6. Contrast ratios of lines to backgrounds.

A second factor favouring contrast is the density of the discrete energy image as compared with that of the background. In order to examine this second factor, the number of additional focused β-rays was taken as 0·2%, 1% and 2% of the total number present in the background (still maintained uniform and lying between 80% and 120% of the focused ($H\rho$)) and the new image densities were determined. Hence the ratio of the maximum central density of the combined image of line plus background to that of background alone can be calculated to give the contrast of the focused image over the background line. The graphs of fig. 6 show the results. It is obvious from them that for appreciable contrast, the monokinetic focused β-rays must have at least 2% of the integrated background intensity and that a source width of less than 0·004 R is desirable.

(*a*) *Constancy of i/θ.* The fundamental equation (7) implies that for given $(H\rho)$, i/θ is a constant for a pair of conjugate planes. But eqn. (7) was derived under paraxial conditions, and it is thus extremely important to be able to estimate the accuracy of the equations when they are applied to a real lens working under non-paraxial conditions.

If the lens is stopped down until the aberration skirt is removed, then aberrations which are dependent on the lens aperture may be assumed to be absent. This leaves only astigmatism and distortion. At unit magnification the isotropic distortion and the anisotropic astigmatism are zero for a symmetrical lens (Dietrich 1951). In the lens used in the pilot experiment, the isotropic astigmatism was only 1 mm for an object 1 cm off the axis, an amount negligible in practice, since it was found that the current could be changed by amounts up to 0·2% without much change in image definition. The 1 mm of isotropic astigmatism corresponds to a change of current equal to only 0·05%.

The anisotropic distortion which would bend the line was equal to 0·014% of the total rotation for a point 1 cm off the axis. This amount is much too small to affect the value of θ, and thus with the lens stopped down i/θ is a constant. If the lens is not stopped down, then aberrations dependent on the aperture will be present. Calculations show that the main contribution to the maximum intensity of the line image comes from a narrow annulus, rays passing through the mean radius of which are accurately focused in the image plane. Thus if the paraxial focus is selected, the line image will be substantially formed by rays which do not suffer from aberrations, and thus the value of i/θ will be the paraxial one appropriate to formula (7). If, however, another focus is chosen, the rays forming the image will not be free from aberration, and the rotation of the image will not be given by (7).

A lens may thus be used in two ways, either stopped down, when a range of currents is permissible, or unstopped, when it is essential that the paraxial focus be located. The first method is better because it gives a valuable depth of focus, and owing to the absence of aberrations the line is very sharp. The second method is useful if the source intensity is low, and the strength of the line to background is low since there will be a gain in contrast of line over background at some expense of definition, and the exposure times will be shorter than with the first method.

§4. PILOT EXPERIMENT

In order to test the method practically, a long lens consisting of eight layers was constructed, the windings being 30 cm long with a mean diameter of 6·65 cm whilst the inside diameter of the lens tube was about 5 cm. The lens which worked at unit magnification had an object to image distance of 90 cm, and a focal length of about 23·5 cm. The lens was designed to be supplied by a 60-volt accumulator battery and to focus electrons of 2000 gauss cm when carrying a current of about 1·7 amp.

Photographic plates were used to record the image and these were cut to fit the plate holder (see fig. 7) which fitted into the camera box, which in its turn fitted into the lens tube, rubber gaskets providing vacuum tight seals. The camera box was fitted with a light-tight sliding panel which could be opened when required by means of a rod protruding from the case, a vacuum-tight joint being obtained by means of a Wilson seal.

The approximate current (within 1%) necessary to focus electrons of given energy was obtained from calculations performed on the lens data. The thorium F line was selected for study, and for the very rough initial focusing a wire of 0·25 mm diameter was used. Later wires of 0·06 mm diameter were used.

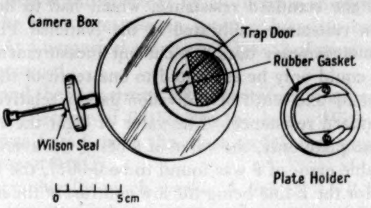


Fig. 7. Camera box and plate holder.

(i) *Axial Astigmatism*

It soon became apparent that the lens suffered from axial astigmatism, and in order to find the orientation of the source so that it lay along one or other of the focal line axes, a protractor scale was attached to the end of the source holder, a suspended plumb line acting as a fiducial line. By rotation of the source holder and refocusing for each orientation the focal line axes were found, the orientation of the source being critical to about ±4°. On reversing the current through the lens windings, very nearly the same orientations for the focal lines were found, the angular difference being about 5° between the two sets. The difference in focusing currents between left-hand and right-hand rotated images with the source orientated along a focal line axis was found to be equal to that calculated to allow for the horizontal component of the earth's magnetic field. As has been mentioned previously, the formula for the rotation of the image in terms of θ, i and n is uniquely true at the paraxial focus, and true at other parts of the focal region provided i is proportional to θ. It was decided in the pilot experiment to use an unstopped lens and to work at the paraxial focus, experiments on stopping down being conducted on another lens. It was not hard to locate the paraxial focus since the line image is thinner and sharper there than at any other part of the focal region. Beyond the paraxial focus the image rapidly disappears while it becomes denser and wider and the edge less well defined as the disc of least confusion is approached. The angle between the two lines obtained by reversing the current was measured on a toolmaker's microscope. The photographic plate was mounted on a glass work-stage and the angle measured by means of a special protractor ocular carrying scales so that direct reading to a minute of arc was possible.

(ii) *Results and Accuracy*

The value of θ for the F line of thorium B was found from five plates, θ being measured ten times for each plate. The mean for each plate was found together with the most probable error. The mean of the five plates was found by weighting each plate proportional to the inverse square of its probable error according to standard practice. This gave a value of 1·40568 radians for θ with a relative probable error of 0·00013 for the arithmetic mean. The total relative probable error was estimated to be 0·0014. This included contributions from the uncertainty in the number of turns in the lens, the uncertainty in the end effect, the uncertainty in the potentiometer and galvanometer readings, and the uncertainty in the standard resistance. Also included was a value for the purely instrumental accuracy of the optical measuring device for θ. The largest error came from the uncertainty in the standard resistance, which had to be compared with a standard 0·01 ohm resistance calibrated at the National Physical Laboratory. To do this the potentiometer used for current measurements in the lens was used, and settings could only be estimated to one tenth of the smallest division. This possible error of one tenth of a division gave a relative probable error of 0·0003 for the standard resistance. The value of θ for the I line of thorium B was found from one plate only, the value of θ being the average of ten readings. The relative probable error of θ was found to be 0·0017, the reason for the large increase over that for the F line being the low contrast of the image and relatively poor definition as compared with the F line, and also the fact that the result was given by only one plate.

The specific momenta in gauss cm are F line: $(H\rho)=1385{\cdot}9\pm2{\cdot}0$, I line: $(H\rho)=1757{\cdot}0\pm4{\cdot}0$.

From the above results it is seen that the error contributed by the uncertainty in θ for the F line is only 1/7500. In a precision experiment based on the results of the pilot experiment alone, the errors contributed by the current and the number of turns would be reduced to the order of 10^{-5} each. This would require a lens having about 3×10^4 turns.

§5. FURTHER EXPERIMENTAL WORK

In view of the great promise of the pilot experiment, further experimental work was carried out with another lens. The purpose of the work was to test the constancy of i/θ and to examine the effect of different lens apertures and different source widths. The lens used was in general similar to the one used in the pilot experiment, and the current measuring equipment was the same: the number of turns was, however, not known exactly as the lens had been designed for another purpose.

(i) *The Influence of the Aperture*

The theoretical section shows that a reduction of aperture produces two mutually conflicting results, one of which is to reduce the aberration skirt and thus improve the line definition, and the other to reduce the contrast between the discrete line and the background line, the overall intensity of the image being diminished. The optimum conditions for operation therefore appear from theoretical considerations to be those for which the aperture size is of such a value that the aberration skirt is just removed and the source width is less than 0·004 R.

Exposures were made with numerical apertures in ratios of $1:\frac{1}{2}:\frac{1}{4}:1/4\sqrt{2}$, the largest having a value of 0·0467. The temperature of the developer was carefully controlled and development times were regulated by stop-watch. Some of the plates were later microphotometered. Exposures were also made with three widths of source; for the middle one (0·025 mm), which was the most used, the ratios of source width to stop radius were 0·00093, 0·0019, 0·0037, and 0·0052 respectively for the four stops.

It was found in the case of the very strong F line of thorium B that the image of the line became narrower as the numerical aperture was reduced from the maximum to one quarter of the maximum. Use of the smallest stop produced practically no further decrease in image width. There was a slight but not serious decrease in the density of the line image at its centre as the aperture was reduced to one quarter of its maximum. At this last aperture equality of density was attained with that at full aperture by an increase in exposure time of about 50% in spite of the total transmission being reduced to one sixteenth of that at full aperture. At a numerical aperture of $1/4\sqrt{2}$ of maximum the fall in peak intensity was pronounced.

With this lens and the very strong F line therefore, it appeared that a stop size giving a numerical aperture of 0·012 was the best. With a source width of 0·025 mm which was found to be the most useful, this gives a ratio of source width to tube radius of 0·0037.

For the photographically much weaker A line of Th B however, the image obtained with the lens working at a numerical aperture of 0·012 was much less distinguishable from its background than it was when taken with the lens at full aperture. In this case the initial line intensity to the initial background intensity is so low that the fall in contrast due to the increased ratio of source width to tube radius in stopping down (see fig. 6) must be more important than improvement in the fineness of the image by removal of the spherical aberration skirt.

As the purpose of the two line images is to obtain the angle between them with the greatest possible precision, it is obvious that in stopping down some balance must be struck between image contrast and image fineness. Increased line fineness is certainly desirable when possible. If the line to background contrast is low however, and the lens is used unstopped so as to improve the line image to background contrast, then it is essential that the image be at the paraxial focus in order that the fundamental equation (7) may yield the correct value of $(H\rho)$ (see § 3 ii (*a*)). In the particular case of the F line images, the use of a quarter aperture stop (when the spherical aberration skirt was just removed) as against full aperture enabled the precision with which the angle between the lines could be measured to be improved by a factor of nearly 3. This reduced the probable error of the mean of ten measures to rather less than ± 0·2 minute or rather less than 1/20 000 of the total rotation.

(ii) *The Constancy of i/θ*

With full aperture (N.A. 0·047) it was found that the instrument had an appreciable depth of focus, the total range being about ⅛% of the focusing current. As the current was varied between extremes of this range no detectable change in the rotation could be observed so that at full aperture the determination of the paraxial focus was essential. On stopping down to quarter aperture, however, it was found that current and rotation became exactly proportional.

Plates were taken at each of five values of focusing current in a range of ⅓% of the focusing value, this increased range being easily possible at quarter aperture. The five means did not differ in extreme by more than 1/3000 and showed no trend with increasing current. Sixteen plates were taken in all and ten measures of θ for each plate gave 160 determinations of the ratio of i to θ. Normal distribution of these about a mean has been demonstrated, the probable error of any one determination of i/θ being $\pm 0{\cdot}024\%$ of the mean value.

(iii) *The Influence of Source Width*

The effect of three source widths of 0·100, 0·025 and 0·010 mm respectively was tried. Two important conclusions follow: the first is that the exposure time necessary for an image of satisfactory density need not be increased at all as the source width is reduced from 0·100 mm to 0·01 mm; the second is that the line image becomes finer in proportion as the source width is changed from 0·100 mm to 0·025 mm, though beyond this the decrease in image width is only slight. Possible reasons for this are: slight unsteadiness in focusing currents and currents in degaussing frame, the grain size of the emulsion, and the presence of small uncompensated stray fields.

§6. CONCLUSIONS

From the results of the experimental work with the second lens, it was found that with a source width of 0·025 mm and a numerical aperture of 0·0123, and taking i/θ as the mean of fifteen plates, i/θ was known with a relative probable error of less than $1/(1{\cdot}5\times10^4)$. As the measurement of the current and the number of turns can be made with an accuracy approaching $1/10^5$ it is seen that $(H\rho)$ can be found even with the present apparatus to better than $1/10^4$. It is proposed to build a lens with about 3×10^4 turns and to use precision current measuring equipment, together with a new optical device for measuring θ. With this new optical device a probable error of about 0·2 minute for any one reading is expected. The probable error of 25 readings would thus be of the order of 0·04 minute, giving in the case of the Th F line a relative probable error of about $1/10^5$ for θ. Using the accurate current measuring equipment the relative probable error in $(H\rho)$ should thus be better than $2/10^5$ for the thorium F line.

PART II

§1. INTRODUCTION

The observations described here were made with a lens on which the exact number of turns was not known. As explained in §§1, 2 and 3 of Part I, such a lens cannot be used for absolute determinations of β-ray $(H\rho)$ but still enables comparisons to be made. For this purpose only the ratio of the mean focusing current to the angle between the focused lines must be ascertained for each of the lines being examined and the quotients are then in the ratio of the corresponding β-ray $(H\rho)$.

Photographs were taken of the I and F lines of thorium B and the A line of thorium C. Of these, the two first-named β-ray lines arise in different electronic shells excited by the same γ-ray, and it has been pointed out by Siegbahn (1941) that in such a case a comparison between the $(H\rho)$, coupled with the difference in binding energy of the shells concerned (obtained accurately from x-ray data), is sufficient to provide the absolute values of the β-ray $(H\rho)$. All three lines, being particularly strong against the continuous spectra accompanying them, are especially suited to the method.

§2. THE MEASUREMENT OF i

The actual measurements made were of the potential difference in volts, measured with a potentiometer, across the terminals of a standard resistance carrying the lens currents. A constant of proportionality would supply the values of these currents, but as the apparatus was used for $(H\rho)$ comparisons only, the magnitude of this constant was not required. All currents were well within the limiting value for the resistance, and the ambient temperature did not fluctuate by more than ±1°C. Calibration of the potentiometer by the National Physical Laboratory gave its readings correct to ±1 in 18000, ±1 in 14000 and ±1 in 5400 respectively for the I, F and A lines. The increasing possible error in going from the I line to the A line is due to the decreasing currents required to focus them.

The systematic probable error in the measurement of i due to the instrument was taken as one half the maximum possible, viz. ±1 in 30000 for the I and F lines and ±1 in 10000 for the A line.

§ 3. THE MEASUREMENT OF θ

The angle between the crossed lines was measured on a toolmaker's microscope fitted with a protractor ocular. On an average this angle lay within a few minutes of 73·5°: the sum of the two currents used to focus the lines was constant to within ±0·4%, and i/θ is a constant, so that the maximum variation of θ from 73·5° was ±10′. By arranging every plate so that the scale reading for the first line lay near to the centre of the first degree and measuring the acute angle between the two lines, the second scale reading invariably lay between 33° and 34°. From these two readings the value of θ is calculated; as the determinations of i/θ were made only for comparison, any small systematic errors in the scale of the instrument affected all results in the same way and could therefore be neglected.

§ 4. i/θ FOR THE F AND I LINES

Sixteen photographs were taken of the F line over a range in mean focusing current of 0·3%. Ten measures of θ were made on each photograph making in all 160 determinations of i/θ, of which the mean was $16048{\cdot}5 \times 10^{-9}$ volt; the residual R and the square of the residual R^2 could be obtained for each value of i/θ. To show the spread in results the mean value of i/θ for each of the 16 plates is given in table 2 together with the residual for each. In order to demonstrate normal distribution of these results, and therefore constancy of i/θ with varying θ, they were plotted as a histogram, and compared with the appropriate gaussian distribution curve having the general equation $y = nh\pi^{-1/2}\exp(-h^2x^2)$. In this h the constant of precision which controls the shape of the curve can be obtained either from the residuals or the residual squares. The total area under the curve gives the number of determinations n of i/θ. The two values of h so obtained were the same, an indication of how closely the determinations of i/θ approached normal distribution, whilst the probable error in any one reading, obtained from $\pm 0{\cdot}477/h$, was $\pm 3{\cdot}8 \times 10^{-9}$. In obtaining the probable error in the mean of all the values of i/θ account was taken of the fact that there were 16 independent readings and not 160, giving a probable error in the mean of $\pm 0{\cdot}95 \times 10^{-9}$ with an uncertainty of about 12%. Combining this with the error in i we find a final value of i/θ for the F line of $(16048{\cdot}5 \pm 1{\cdot}1) \times 10^{-9}$.

Table 2. Mean Values of i/θ for Observations of the F Line

$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$
16047·0	1·5	16043·6	4·9	16047·7	0·8	16044·5	4·0
16054·0	5·5	16047·8	0·7	16048·3	0·2	16039·8	8·7
16052·5	4·0	16052·2	3·7	16051·7	3·2	16054·9	6·4
16040·3	8·2	16057·0	8·5	16048·5	0·0	16045·8	2·7

An identical process was performed on the 90 values of i/θ obtained from nine plates of the I line. As with the F line, gaussian distribution was confirmed and the final value of i/θ for the I line was obtained as $(20\,272{\cdot}0 \pm 1{\cdot}6) \times 10^{-9}$ volt. Mean values and residuals for each plate are shown in table 3.

Table 3. Mean Values of i/θ for Observations of the I Line

$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$
20265·1	6·9	20275·7	3·5	20275·7	3·7	20274·7	2·7
20266·0	6·0	20266·4	5·6	20268·0	4·0	20274·3	2·3
20282·7	10·7						

§ 5. THE ABSOLUTE VALUES OF THE I LINE AND F LINE ($H\rho$)

$(H\rho)_{\mathrm{I}}$ for the I line is related to $(i/\theta)_{\mathrm{I}}$ for the same line by the equation $(H\rho)_{\mathrm{I}} = p(i/\theta)_{\mathrm{I}}$ where p is the calibration constant of the lens. The voltage V_{I} required to impart to an electron a specific momentum $(H\rho)_{\mathrm{I}}$ is given by the equation

$$V_{\mathrm{I}} = 10^{-8}\frac{m_0\mathbf{c}^2}{\mathbf{e}}\left\{\left(1+\left[\frac{\mathbf{e}}{m_0\mathbf{c}}\,p\left(\frac{i}{\theta}\right)_{\mathrm{I}}\right]^2\right)^{1/2}-1\right\}.$$

Similar equations apply between the F line values. If we know $V_{\mathrm{I}} - V_{\mathrm{F}}$, the difference between the binding energies of the K and L shells of element 83 in which the F and I lines respectively arise, we can solve the equations for p. Bergstrand's (1949, 1950) value of $299\,793{\cdot}1 \pm 0{\cdot}25$ km/sec was taken for $\mathbf{c}$ and Thomas, Driscoll and Hipple's (1950) value of $(1{\cdot}75890 \pm 0{\cdot}00005) \times 10^7$ e.m.u./g taken for $\mathbf{e}/m_0$. The value of $V_{\mathrm{I}} - V_{\mathrm{F}}$ was given by Mlle. Y. Cauchois (private communication) as $74\,128 \pm 7$ electron volts.

This gave $p = (8{\cdot}6520 \pm 0{\cdot}0017) \times 10^7$ from which the I line and the F line $(H\rho)$ were found to be $1753{\cdot}94 \pm 0{\cdot}39$ and $1388{\cdot}52 \pm 0{\cdot}30$ gauss cm respectively.

These values coincide exactly with those of Lindström (1951 a) determined with very high precision using a semicircular focuser. It is interesting to note that of the earlier measurements, those of Surugue (1937) which were the most precise, also coincide most nearly with these very exact later ones.

§ 6. RESULTS FOR THE A LINE

The contrast of the A line to its accompanying background was found to be rather poorer than that of either the I line or the F line, causing a slightly greater spread in the determinations of θ on any one plate. The number of photographs taken was therefore increased to 14. Results for the means are given in table 4.

Table 4. Mean Values of i/θ for Observations of the A Line

$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$	$i/\theta \times 10^9$	$R \times 10^9$
6166·98	1·16	6166·05	2·09	6168·11	0·03	6167·33	0·81
6171·58	3·44	6169·06	0·92	6171·22	3·08	6165·44	2·70
6166·54	1·60	6168·78	0·64	6169·84	1·70	6165·48	2·66
6167·10	1·04			6170·40	2·26		

The arithmetic mean of the 140 determinations of i/θ for the A line was found to be $(6168{\cdot}1 \pm 0{\cdot}8) \times 10^{-9}$ volt. This, combined with the value of the calibration constant of the lens gives the A line $(H\rho)$ as $(533{\cdot}66 \pm 0{\cdot}12)$ gauss cm, corresponding with Surugue's less precise determination of 534. Arnoult (1939) also obtained 534 but did so by ratio with the $(H\rho)$ of the F line; as his F line value was appreciably lower than either Surugue's or the result obtained in this work, the correspondence has little significance. Of the other investigators who made observations on the A line, Flammersfeld (1939) obtained a value of 536·5 while Ellis (1932) and Sze (1933) obtained rather higher values of 537 and 541 respectively. Furthermore, except for Surugue, who obtained 0·3846, all had ratios of A line $(H\rho)$ to F line $(H\rho)$ nearly $\frac{1}{4}\%$ higher than the value 0·38434 found here. It is thought that the precision which has been attained in the measurements outlined here is greater than that of the above workers, and that the $(H\rho)$ value of the A line has now been established more accurately than has so far been achieved.

ACKNOWLEDGMENTS

This work formed part of a programme of research, supported by a grant from the Department of Scientific and Industrial Research, and directed towards the improvement of β-spectroscopy.

We would like to thank Professor J. D. Bernal for allowing us the facilities of his laboratory and Mr. R. E. Siday for suggesting the problem, and for discussions arising as the work developed. Thanks are also due to Dr. J. C. E. Jennings for helpful discussions.

REFERENCES

ARNOULT, R., 1939, *Ann. Phys., Paris*, **12**, 241.
BERGSTRAND, E., 1949, *Nature, Lond.*, **163**, 338; 1950, *Ark. Fys.*, **2**, 15.
CRAIG, H., 1952, *Thesis*, University of London.
DIETRICH, C. F., 1951, *Thesis*, University of London.
ELLIS, C. D., 1932, *Proc. Roy. Soc.* A, **138**, 18.
FLAMMERSFELD, A., 1939, *Z. Phys.*, **112**, 727, **114**, 227.
LINDSTRÖM, G., 1951 a, *Phys. Rev.*, **83**, 465; 1951 b, *Ark. Fys.*, **4**, 1.
MALOFF, I. G., and EPSTEIN, D. W., 1938, *Electron Optics in Television* (New York: McGraw-Hill.
ROGERS, F. T., Jnr., 1936, *Phys. Rev.*, **50**, 515.
SIEGBAHN, K., 1941, *Ark. Mat. Astr. Fys.*, **28**, 6; 1944, *Ibid.*, **30**, 20.
SCOTT, F. A., 1934, *Phys. Rev.*, **46**, 633.
SIDAY, R. E., 1942, *Proc. Phys. Soc.*, **54**, 266.
SURUGUE, J., 1937, *Ann. Phys., Paris*, **8**, 484.
SZE, S.-Y., 1933, *Ann. Phys., Paris*, **19**, 59.
THOMAS, H. A., DRISCOLL, R. L., and HIPPLE, J. A., 1950, *Phys. Rev.*, **78**, 787.

Transistor Action and Related Phenomena in Lead Sulphide Specimens from Various Sources

By C. A. HOGARTH
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 22nd October 1952

Abstract. Examination of a large number of specimens of PbS has indicated a maximum free carrier concentration at 290°K of $2 \times 10^{17}\,\mathrm{cm}^{-3}$ if transistor action is to be observed. This figure leads to a value of 0·65 ev for the width of forbidden band at 290°K. Transistor action is only observed in p-type specimens and is always associated with good rectification and a strong photo-voltaic effect.

---

§ 1. INTRODUCTION

TRANSISTOR action in PbS has been described by Banbury, Gebbie and Hogarth (1951) and results quoted for assemblies made with crystals taken from Zeiss photo-voltaic cells. These crystals were of Sardinian galena which had been specially treated by optically polishing and warming, before storing *in vacuo*. Little information was given about the conduction properties of the crystals except that for one of them the value of the conductivity was about $1\,\mathrm{ohm}^{-1}\,\mathrm{cm}^{-1}$ and the hole mobility about $350\,\mathrm{cm}^2$/volt second. Since that time Putley and Arthur (1951) have measured the conductivity and Hall constant of these crystals as functions of temperature and have found that above 900°K PbS acts as an intrinsic semiconductor with a forbidden band width of 1·17 ev. Furthermore the forbidden band width is found to decrease with temperature. The value at 290°K is of great importance in any attempt to explain barrier layer phenomena involving PbS and various different methods of computing this quantity are available. From the measurements to be reported in this paper, a value has been estimated and this agrees with the results of other workers.

The results of measurements on many PbS specimens indicate a limit to the free carrier concentration at 290°K beyond which transistor action is not expected to occur. Furthermore they show that PbS crystals from several sources are suitable for experiments on transistor action.

§ 2. EXPERIMENTAL METHODS

The various point contact properties were examined for etched or cleaved surfaces of PbS crystals taken from various sources. The rectification and transistor action were examined by means of simple test circuits and oscilloscope displays in conjunction with a micro-manipulator for the adjustment of the whisker contacts. The photo-voltaic effect was tested with a fine spot of white light and a whisker contact, the signal being measured by a galvanometer, or by an amplifier and ammeter when interrupted light was employed. Values of Hall constant R and electrical conductivity σ were measured by standard d.c. potentiometer methods.

§ 3. Experimental Results

Experiments were carried out on a large number of crystals from various sources and the table includes results for nineteen representative specimens. (Some of the measurements were carried out at Reading University, the others at Telecommunications Research Establishment. The measurements of R and σ for specimen X 1 were made by Dr. E. H. Putley.) It may be noted immediately that transistor action is observed in specimens of PbS from various sources and although Sardinian galena can show transistor action, material from other sources may prove superior.

The main features of the results may be summarized as follows: (i) Good rectification, photo-voltaic effects and transistor action are only observed in p-type specimens of high purity as indicated by a large Hall constant and high carrier mobility. (ii) Electroforming effects are most marked in p-type specimens of high purity but are also observed in n-type specimens which have a large value of Hall constant. (iii) Highly impure n-type specimens show no sign of the existence of rectifying barriers. (iv) N-type specimens of high purity do not exhibit transistor action in spite of values of R and σ comparable to those p-type specimens which do so.

§ 4. Discussion of Results

In the absence of impurity scattering and mixed conduction the product $R\sigma$ with a small correcting factor (approximately unity) would yield a direct measure of carrier mobility. Detailed information on impurity scattering in PbS is not available, but according to E. H. Putley (private communication) it may be neglected. Since the measurements were carried out at 290°K and intrinsic conduction is not expected to occur at temperatures as low as this it is unlikely that mixed conduction is occurring. Thus we may take the product $R\sigma$ as an approximate estimate of mobility. It should be noted that the mobilities quoted for p-type specimens refer to holes. The injected electrons concerned in transistor action would be expected to have even higher mobilities. Although a high carrier mobility is usually considered necessary for a specimen to exhibit transistor action, this is not a *sufficient* condition as may be observed for specimens N9 and A2 (see table). A fairly large value of R and a low value of σ are also necessary and for transistor action to occur in PbS it would seem that the concentration of free holes at room temperature should not be greater than $2\times10^{17}\,\mathrm{cm}^{-3}$.

It was pointed out by A. F. Gibson (private communication) that this figure might enable the width of forbidden band in PbS at 290°K to be estimated. The calculation was carried out as follows: Let ϕ be the height of the surface barrier, E the width of forbidden band and ϵ the distance from the Fermi level to the top of the full band (see figure). Then to achieve barrier conditions favourable for electron injection we use the condition given by Bardeen and Brattain (1949), namely $\phi+\epsilon\geqslant\frac{1}{2}E$, and in the limit we take the equality. The average value of ϕ for PbS derived from measurements of rectification, photoconductivity and photovoltage made by various authors may be taken as 0·2 ev. Furthermore the activation energy for impurities $\Delta\epsilon$ has been found to be of order $\boldsymbol{k}T$, i.e. 0·025 ev. Then the required value of ϵ is given by $\eta\boldsymbol{k}T$ where η is the reduced electrochemical potential for holes. Lehovec and Kedesdy (1951)

Summarized Properties of various Lead Sulphide Specimens

Specimen No.	Origin	R (cm^3/coul)	σ (ohm^{-1}cm^{-1})	Type	$R\sigma$ (cm^2/vsec)	Carrier† Concentration cm$^{-3}\times 10^{16}$	Rectification	Photo-voltaic Effect	Transistor Action
Z3	Zeiss Ikon Cell (Sardinia)	+198	2·06	P	410	3·7	Good*	Strong	Yes
Z4	,,	+170	4·20	P	710	4·3	Good*	Strong	Yes
N1	British Drug Houses (Sardinia)	−144	9·31	N	1340	5·1	Poor*	Weak	No
N2	,,	+347	0·81	P	280	2·1	Good*	Strong	Yes
N3	,,	−1·42	17·2	N	24	520	Poor	Weak	No
N5	,,	+42	1·24	P	52	18	Good*	Weak	Yes
N9	,,	+3·6	74	P	270	200	Poor	Weak	No
G3	Gregory Bottley (Sardinia)	+102	2·12	P	220	7·3	Good*	Strong	Yes
S3	Sardinia (Miniere di Monte Vecchio)	−2·5	330	N	830	290	Poor	Weak	No
S16	,,	−139	1·52	N	210	5·3	Poor*	Weak	No
I2	Ireland	−0·85	750	N	640	940	None	None	No
C3	Cumberland	−3·46	46	N	160	210	Poor	Weak	No
X1	Saxony	+50	1·2	P	60	1·5	Good*	Strong	Yes
W1	Wisconsin	−110	1·0	?	110	6·8	Good*	Strong	Yes
J1	Joplin	+52·5	0·95	P	50	14	Good*	Strong	Yes
J3	,,	+200	1·21	P	240	3·7	Good*	Strong	Yes
U1	Urals	−12·5	21	N	260	60	Poor	Weak	No
A2	Artificial (‡) (T.R.E. Polycrystalline)	+0·31	470	P	150	2400	Poor	Weak	No
A3	Artificial (‡) (T.R.E. single crystal)	+2·4	270	P	650	310	Poor	Weak	No

* Signifies that contact may be formed electrically.

† Based on Hall effect data and uncorrected for impurity scattering and possible mixed conduction.

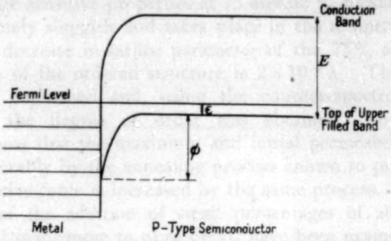
‡ Prepared by W. D. Lawson.

have plotted graphs relating the quantities x and y for various fixed values of η:

$$x=\frac{\Delta\epsilon}{kT}, \quad y=\frac{N_0}{2(2\pi m^*kT/h^2)^{3/2}}, \quad \text{where } N_0=2\times10^{17}\,\text{cm}^{-3}, \quad T=290^\circ\text{K}.$$

Thus for PbS at room temperature $x=1$ and $y=8{\cdot}4\times10^{-3}$. The corresponding value of η is 5 and thus $\epsilon=0{\cdot}125$ ev. Hence $E=2(\phi+\epsilon)=0{\cdot}65$ ev. Calculations made for other likely values of $\Delta\epsilon$ and ϕ suggest a probable error of $\pm0{\cdot}1$ ev in the above value of E.

It is of interest to compare this value with other determinations of the same quantity. By extrapolating Putley and Arthur's curve of $E/2$ against T to 290°K a value of 0·7 ev for the band width at that temperature was obtained. A derivation of E using results obtained with a p–n junction in a PbS crystal (A. F. Gibson, private communication) gave a value of 0·68 ev. [The p–n junction from which this value was obtained was the only PbS p–n junction ever achieved by Gibson having rectification characteristics in agreement with the theory due to Shockley (1949).] Thus within the accuracy of experiment consistent results for E are obtained from three different experiments.



Energy level diagram for contact between metal and p-type semiconductor.

In the case of Wisconsin galena an interesting problem arises. It was often found that specimens (e.g. W1) exhibit a large negative value of R indicating excess conduction whereas the point contact properties are characteristic of a deficit semiconductor. This cannot be explained by supposing the specimen to be in the intrinsic range, since if this were so, good rectification and transistor action would not be expected to occur. It is hoped that experiments on transistor phenomena as functions of temperature will help solve this problem.

Probably the most consistently satisfactory material for the investigation of the transistor effect in PbS is galena from Joplin, Missouri. The specimens examined appear to possess excellent cleavage surfaces, and their contact properties are substantially uniform. Sardinian galena, although occasional samples are of excellent quality, is much less homogeneous, is polycrystalline, and its electrical properties vary rapidly from one region of a crystal to another.

ACKNOWLEDGMENTS

Some of the above experiments were carried out at Reading University under an Admiralty contract and I should like to thank Professor R. W. Ditchburn for the provision of laboratory facilities. For recent discussions and suggestions I should like to thank Dr. E. H. Putley and Dr. A. F. Gibson. Some of the crystals of Sardinian galena employed were kindly provided by the Director of the Monte Vecchio Mine. I am indebted to the Chief Scientist, Ministry of Supply, the Admiralty, and the Controller, H.M. Stationery Office, for permission to publish this work.

REFERENCES

BANBURY, P. C., GEBBIE, H. A., and HOGARTH, C. A., 1951, *Semi-Conducting Materials* ed. H. K. Henisch (London: Butterworths Scientific Publications), p. 78.
BARDEEN, J., and BRATTAIN, W. H., 1949, *Phys. Rev.*, **75,** 1208.
LEHOVEC, K., and KEDESDY, H., 1951, *J. Appl. Phys.*, **22,** 65.
PUTLEY, E. H., and ARTHUR, J. B., 1951, *Proc. Phys. Soc.* B, **64,** 616.
SHOCKLEY, W., 1949, *Bell Syst. Tech. J.*, **28,** 435.

A Study of the Order–Disorder Transformation in Iron–Nickel Alloys in the Region $FeNi_3$*

By R. J. WAKELIN† and E. L. YATES
Physics Department, University of Sheffield

MS. received 26th May 1952, and in final form 15th September 1952

Abstract. The order–disorder transformation in iron–nickel alloys of high purity has been studied by measurements of electrical resistivity, by x-ray analysis and by an examination of the magnetic properties in both low and high fields. It is shown that the transformation appears to take place over a wide range of composition, from about 50 to 80 atomic % nickel, with a maximum change in structure sensitive properties at 75 atomic % nickel. The formation of order is extremely sluggish and takes place in the temperature range 494°c to 500°c. The decrease in lattice parameter of the 75% alloy which occurs on the formation of the ordered structure is 2×10^{-3} Å. The existence of the superlattice was confirmed and, using the counter-spectrometer technique, an estimate of the degree of order was obtained. Low field magnetic measurements show that the maximum and initial permeabilities of $FeNi_3$ are decreased considerably by the annealing process known to produce the ordered alloy. The coercive force is increased by the same process.

The effects of the addition of small percentages of aluminium, copper, molybdenum and manganese to pure $FeNi_3$ have been examined. It is shown that the experimental results support the assumption that the 'spare' electrons of the added elements enter the 3d shell of the nickel atoms only, the contribution of the iron atoms to the total magnetic moment being unaltered Thus the iron and nickel atoms may be considered to act independently.

§ 1. Introduction

The elucidation of the widely varying properties of iron–nickel alloys has been the subject of numerous researches. On account of its magnetic properties, the region extending from 35% to 90% nickel, the 'Permalloy' region, has engaged considerable attention, but the physical interpretation of the collected results presents great difficulty. Dahl, in 1936, first suggested the possibility of an ordered structure occurring at 75 atomic % nickel. Numerous attempts have been made to substantiate this suggestion, with varying degrees of success. Leech and Sykes (1939), in co-operation with Jones and Sucksmith, confirmed the existence of the superlattice, on the basis of specific heat, x-ray and magnetic measurements. The purpose of the work described in this paper was to investigate the effect of the degree of order on the magnetic and other relevant physical properties and, as far as possible, to correlate them with associated physical phenomena. In order to eliminate possible complications due to the presence of impurities, alloys were melted under the best possible conditions from metals of the highest purity obtainable. The conditions under which the order–disorder transformation takes place have been examined by electrical resistivity measurements, by x-ray analysis, and by magnetic measurements made on the alloys in high and low fields.

---

* Part of a thesis submitted by one of us (R. J. W.) and approved by the University of London for the external degree of Doctor of Philosophy.

† Now at G.K.N. Research Laboratories, Ettingshall, Wolverhampton.

§ 2. THE PREPARATION AND HEAT TREATMENT OF THE ALLOYS

The alloys were prepared from high purity metals by melting *in vacuo* in a high-frequency induction furnace. Small pieces of electrolytic iron of spectroscopic purity containing about 0·7% oxygen were mixed in the required proportions with nickel pellets of 99·94% purity in pure alumina crucibles. By melting rapidly under a pressure too low for an electrodeless discharge to occur, the high purity of the constituent metals was maintained in the resulting alloys. Composition and purity were confirmed by chemical analysis; the total impurity in any one of these alloys was less than 0·2%. About 50 alloys were prepared in this way, ranging in composition from 45 to 85 atomic % nickel, and also a series of alloys of composition $FeNi_3$ plus small proportions of aluminium, molybdenum, copper and manganese.

Each alloy was carefully hot forged and then homogenized at 1200°c for 18 hours. Photomicrographs showed that this treatment was sufficient to ensure uniform composition throughout the mass of the alloy, and thus experimental evidence obtained from a small part of the material was representative of the behaviour of the complete alloy.

The ordering process may, in general, be accomplished either by annealing for some time at a steady temperature below the critical temperature or by cooling at a fixed rate through a wide temperature range, for example from just above to well below the critical temperature. Even if, at the commencement of this work, the critical temperatures of the alloys had been known, comparison between them would have been difficult if they had not been annealed under strictly comparable conditions. Preliminary measurements showed that the transformation took place over a range of temperature, and it was therefore considered advisable to anneal all the alloys similarly by cooling slowly over a wide temperature range.

It has long been known that the approach to equilibrium in iron–nickel alloys is very slow. It follows that ordering in alloys with a low critical temperature will be still more sluggish, and thus tend to show a less degree of order for a given rate of cooling. As will be shown later, extremely slow rates were used in order to minimize this effect. The range of temperature over which slow cooling was necessary was found from early resistivity measurements. When an alloy is cooled slowly, two processes proceed simultaneously: (*a*) certain nuclei tend to grow at the expense of others and (*b*) the degree of order in each nucleus tends to increase as the temperature is reduced.

Disorder may, in general, be obtained either by cooling rapidly from a temperature above the critical temperature, or by performing cold work on an ordered alloy. The former method only was used.

§ 3. ELECTRICAL RESISTIVITY MEASUREMENTS

(i) *Experimental*

To attain accuracy and reproducibility in the resistance measurements it was desirable to use strips of alloy about 15 cm long and of small cross section. This requirement proved difficult to attain with the high purity alloys as they were brittle and frequently fractured during machining. However, the forging and homogenizing treatment greatly eased the task of machining thin strips about 10 cm long, 4 mm wide and 1 mm thick. The cross section was uniform to within ± 0·02 mm in each direction.

The resistance of the alloy strip was found by the standard technique of comparing the potential difference across a measured length of the strip with that across a standard resistance through which the same current flowed. The use of nickel knife-edges as potential and current terminals enabled the effects of thermal e.m.fs. to be reduced. The corrections for such e.m.fs. were never more than 0·1% of the potential being measured. The reproducibility of the results showed that the resistivities of the alloys were measured to better than ± ½% at any temperature up to about 600° c. Resistance measurements were not made above that temperature, it being unnecessary to do so with the alloys under examination.

Careful control of the temperature in all furnaces was essential. The temperature controller described by Yates (1946) was found to be very efficient and easily adapted to control a continuous variation of temperature. When used with a suitable furnace circuit the controller kept the temperature constant to within ± 0·5° c at all temperatures used. To cool a furnace slowly it was simply necessary to connect a variable resistance in series with the platinum resistance thermometer and to increase this resistance at a steady rate.

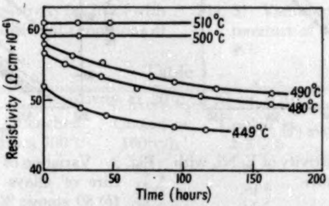


Fig. 1. Variation of resistivity with time at fixed temperatures (pure $FeNi_3$).

Preliminary work on alloys of commercial purity had indicated the approximate critical temperature of the alloy $FeNi_3$. Further preliminary experiments had also shown that the cooling rate must be extremely slow (less than 0·3° c per hour) if a high degree of order is to be obtained. The alloys exhibit considerable thermal 'hysteresis'. The upper temperature limit, above which annealing to produce order is ineffective, was found to be about 500° c ; the lower temperature limit, below which the transformation is too sluggish for any appreciable degree of order to form in a sensible time, was about 400° c. A wide range of resistivity may be obtained at any one temperature below 530° c, depending on the previous heat treatment and the length of time the alloy is held at that temperature.

The alternative method of producing order, namely by annealing the alloy for a given length of time at a constant temperature, was tried with a specimen of the pure material. The changes in the resistivity of the disordered alloy with time, at certain constant temperatures, are shown in fig. 1. Before each set of measurements the alloy was disordered by heating to 600° c, annealed for 12 hours and then rapidly cooled to the temperature at which the investigation was to take place. These results confirm the earlier conclusion that the 'critical temperature' lies in the region 490° to 500°c. The curves show clearly the slow rate at which the transformation proceeds, an equilibrium degree of order not being obtained at any temperature even after annealing periods of more than 200 hours.

A few resistance measurements were made at low temperatures. Extrapolation to absolute zero from the temperature of liquid oxygen suggested that the alloys were not fully ordered even after the longest annealing periods given, since the resistivity of a perfectly ordered lattice at the absolute zero should, theoretically, be zero.

(ii) *The Effect of Composition on the Formation of Order*

Alloys in the composition range 70 to 80 atomic % nickel were prepared and the resistivities measured as already described. Some of the results are illustrated in figs. 2 and 3. Several interesting features may be seen from these curves. At the slowest cooling rate (0·10°c per hour), the specimens took seven weeks to cool through the desired range, and even these conditions were insufficient to produce the completely ordered state. The same rate of cooling in the case of the 80% alloy, however, showed little improvement over the faster rate of cooling. In fact equilibrium had apparently been reached down to 475°c by cooling at 1·3°c per hour. It follows from Matthiessen's rule that the slope of the resistivity–temperature curve should be the same for both ordered and disordered alloys provided no transformation is actually taking place. The experimental curves provide excellent support for this reasoning. At temperatures below 450°c these transformations become very sluggish indeed and the resistance–temperature curves for different degrees of order are parallel. The curves also show that order may be developed outside the composition range 70–80 atomic % nickel and that quite appreciably more order is developed in the 70% than in the 80% alloy.

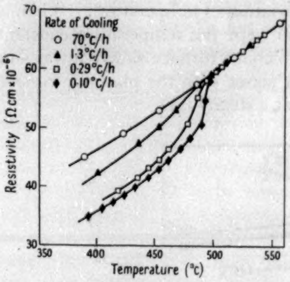


Fig. 2. Variation of resistivity of $FeNi_3$ with temperature.

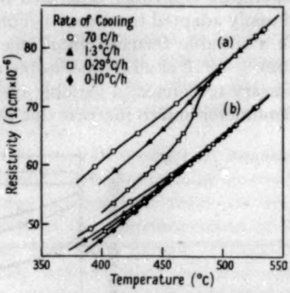


Fig. 3. Variation of resistivity with temperature of alloys containing (*a*) 70 and (*b*) 80 atomic % nickel.

The variation of resistivity with nickel content is shown in fig. 4 after five different conditions of heat treatment. It is seen that some degree of long range order exists at less than 50% nickel but that there is practically none at 82%

nickel. The maximum decrease in resistivity occurs at 75 atomic % Ni, the resistivity of the ordered alloy being 54% of that of the disordered material. This graph is similar to, but not identical with, one given by Kaya (1938).

The difference between the resistivities of the quenched and furnace-cooled alloys may be due to two causes : (*a*) the setting up of strains in the lattice by the drastic quenching process, causing an increase in the resistivity, (*b*) the formation of short-range order in the furnace-cooled alloys.

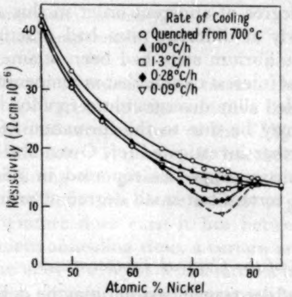


Fig. 4. Variation of resistivity (at 20°c) with composition after different conditions of heat treatment.

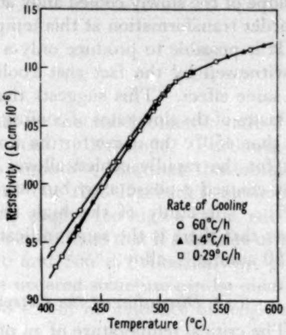


Fig. 5. Variation of resistivity with temperature of an alloy with 50 atomic % Ni.

Table 1

Resistivity at 20°c in Ω cm $\times 10^{-6}$

Composition (atomic %)	Quenched from 700°c	Cooled at 100°c/h	Cooled at 1·3°c/h	Cooled at 0·28°c/h	Cooled at 0·09°c/h
45	42·6	41·4	40·7	40·6	
50	33·1	31·6	30·4	30·3	
55	27·4	25·4	24·6	24·6	
60	23·3	21·3	20·6	20·5	
65	20·0	18·2	17·2	16·9	
68	18·8	16·9	15·7	14·2	
70	18·1	16·3	14·4	12·9	
72	16·7	15·3	13·0	11·4	
74	16·5	15·5	13·0	11·1	
75	16·4	15·0	13·0	11·3	8·86
76	16·2	15·0	13·3	11·5	
78	15·5	14·8	13·8	13·2	
80	15·0	14·2	13·7	13·7	
81	15·1	14·4	14·2	14·2	
85	14·0	13·6	13·5	13·5	

Comparison with previous published work shows that the majority of earlier determinations of resistivity as summarized by Marsh (1938) give values greater than those shown in table 1. This is probably due to the presence of impurities in the alloys used in the earlier work.

There is a reduction in resistivity on prolonged annealing of the 50 atomic % alloy. The existence of a superlattice at 50 atomic % Ni was considered a possibility and attempts were made to test this hypothesis. From theoretical considerations (e.g. Easthope 1937) the critical temperature should be greater than that of the 75% alloy. The 50% alloy was therefore cooled first rapidly and afterwards very slowly from 700°C. No difference was observed for the different rates of cooling at temperatures above 510°C (fig. 5). The Curie temperature of the slowly cooled alloy is about 543°C (table 5). The change of slope of the resistivity curve between 500°C and 540°C is to be related to the change in the magnetic properties. There appears to be an additional change of slope of the slowly cooled alloy at about 505° C which would suggest an order–disorder transformation at this temperature.

It is possible to produce only a small degree of apparent order in this alloy as witnessed by the fact that cooling slowly at different rates had practically the same effect. This suggests that an equilibrium state had been attained in the faster of the slow rates of cooling. It is of interest to note that at temperatures less than 450°C the curve for the slowly cooled alloy diverges more rapidly from that for the rapidly cooled alloys. This may be due to the formation of the body-centred α-phase alloy but was not further investigated (cf. Owen and Liu 1949). The study of the high field magnetic properties reported in § 5 (iii) shows that there is the same indication of a certain but small degree of order in the 50 atomic % alloy.

(iii) *Discussion on the Determination of the Critical Temperatures*

The critical temperature of an order–disorder transformation may be defined as the temperature above which no long-range order can exist in a stable state. In the iron–nickel series of alloys, however, the very sluggish nature of the transformation complicates the determination of such a temperature. Preliminary measurements had shown that it was possible, by rapid heating of an ordered alloy, to retain the ordered state for a while, even 30°C above the critical temperature. The temperature at which the resistance–temperature curve for the ordered alloy meets that for the disordered alloy can be varied by changing the rate of heating.

Källbäck (1947), in an investigation undertaken specifically to determine this temperature, heated an ordered alloy from below the critical temperature and held the temperature steady for long periods at temperatures in the neighbourhood of the critical temperature. He measured the resistivity while keeping the temperature constant, but even after periods of 150 hours the resistivity failed to reach a steady value. He then plotted against temperature the resistivity so obtained (which he assumed to be the equilibrium value) and extrapolated this 'equilibrium curve' to meet the disordered curve at T_c. The value of T_c (506°C) thus obtained is likely to be greater than the critical temperature.

In the present investigation the formation of order was followed while actually taking place. It has been pointed out that heating and cooling, however slowly, produced resistivity–temperature curves showing 'hysteresis' loops. It is clear, therefore, that any degree of order may be produced over a wide range of temperature, depending on how long the alloy has been held at that temperature or, alternatively, how slowly it has been cooled or heated to the temperature concerned. No fixed or definite critical temperature can thus be defined. However, if the earlier definition of the critical temperature is modified to read "that temperature below which some degree of order may be produced and above which the alloy, *if already in the disordered state*, remains in that state", then the temperature at which the resistivity curve for the slowly cooled alloy begins to diverge from the rapidly cooled curve may be taken as the critical temperature. This temperature was determined from the graphs shown in figs. 2 and 3 by plotting the square of resistivity against temperature over the range 490° to 500°C. Defined in the above manner, the critical temperatures of three alloys are shown in table 2.

Table 2

Composition (atomic % Ni)	70	75	80
Critical temperature (°C)	496	498	494

The maximum in the critical temperature at the stoichiometric ratio 3 : 1 does not agree with the tentative theory advanced by Easthope. Furthermore, the critical temperature of the 75% alloy is some 8° C less than that obtained by Källbäck. However, it has been pointed out that Källbäck's method of determining the critical temperature was likely to give too high a value, and the modified definition above is likely to give a lower limit, so that if a definite critical temperature does exist it lies between 498°C and 506°C. Alternatively, given sufficient annealing time, a certain amount of ordered structure can be obtained in the alloy $FeNi_3$ at temperatures below 498°C, but no evidence of long-range order is obtained at temperatures above 506°C.

§ 4. X-RAY ANALYSIS

(i) *Variation of Lattice Parameter with Composition*

Investigations of the iron–nickel system of alloys using the X-ray technique were made by Bradley, Jay and Taylor (1937) and by Owen, Sully and Yates (1937). In both cases, however, the alloys were not in the ordered state ; none of the previous workers was aware of the extremely slow nature of the order–disorder transformation. The lattice parameter of the ordered alloy $FeNi_3$ was first measured by Leech and Sykes (1939) who reported a reduction in the parameter on annealing for 500 hours in the temperature range 490°C to 370°C. By itself such a change in parameter is not sufficient evidence of an order–disorder transformation in a single-phase alloy. Nevertheless, when such a transformation is established from the direct evidence of the superlattice spectrum, the change in lattice parameter may be considered as due to the transformation and the degree of order inferred from the extent of the parameter change. The alloy powders were prepared from the homogenized ingots and were made with a jeweller's saw to avoid risk of contamination from a file. The powder was sieved through a 250-mesh sieve and annealed in evacuated Pyrex glass or silica tubes. To remove the effect of cold work, all powders were given a preliminary anneal of 24 hours at 600° C. They were then either (*a*) quenched in water to maintain the disordered state or (*b*) cooled slowly, at 0·22°C per hour, a process known to be necessary for the production of the ordered state.

The variation of lattice parameter with composition is recorded in table 3. The maximum difference between the parameter of the quenched and slowly cooled alloys occurs at 75 atomic % of nickel. The values for the quenched alloys are in good agreement with those obtained by Owen, Sully and Yates.

It is clear that the change in parameter confirms the earlier statement that such a change may be used as confirmation of an order–disorder transformation. The percentage change in lattice parameter with the development of order is considerably less than the percentage change in resistivity. The latter is a more sensitive property with which to observe the development of order.

Table 3

Composition of alloy (atomic % Ni)	Lattice parameter (Å) at 20°c Quenched (disordered)	Annealed (ordered)
50	$3{\cdot}5865_5$	$3{\cdot}5864_5$
70	$3{\cdot}5606_6$	$3{\cdot}5591_8$
72	$3{\cdot}5583_4$	$3{\cdot}5569_0$
74	$3{\cdot}5557_6$	$3{\cdot}5539_3$
75	$3{\cdot}5544_2$	$3{\cdot}5522_7$
76	$3{\cdot}5534_4$	$3{\cdot}5511_6$
78	$3{\cdot}5505_4$	$3{\cdot}5494_3$
80	$3{\cdot}5480_6$	$3{\cdot}5479_2$

Copper K radiation: $\lambda\alpha_1 = 1{\cdot}54050$ Å, $\lambda\alpha_2 = 1{\cdot}54434$ Å

Following the technique of Leech and Sykes, namely, using monochromatic cobalt Kα radiation, the superlattice in $FeNi_3$ was confirmed. The 50% alloy, however, yielded no additional lines, and the suggestion of an ordered state with this alloy remains unconfirmed. The decrease in lattice parameter obtained on prolonged annealing is the same order of magnitude as the estimated experimental error and hence does not provide support for the evidence of a small degree of order in this alloy.

(ii) *The Effect of the Addition of Aluminium on the Development of Order*

In a thorough and careful investigation of the ternary system Fe–Ni–Al Bradley and Taylor (1938, 1940) reported that when a small amount of aluminium (at least 4%) is added to $FeNi_3$, superlattice lines appear on x-ray photographs. They suggested that the superlattice may be extended to pure $FeNi_3$ by suitable heat treatment below 600°c.

In order to trace the development of the superlattice by the appearance of the additional lines after varying heat treatments, a series of alloys was prepared containing up to 6 atomic % of aluminium in pure $FeNi_3$. With a 9 cm diameter camera the superlattice lines were readily observed in the case of the 6% Al alloy, cooled from 510°c at 7·5°c per hour. With a slower rate of cooling (0·8°c per hour) the superlattice was confirmed with alloys containing as little as 1·5 atomic % of aluminium. It has already been reported in the previous paragraph that the superlattice lines may be obtained with pure iron–nickel alloy provided the cooling rate is slow enough. It follows that the cooling rate of 10°c per hour used by Bradley and Taylor was too rapid to give the equilibrium conditions in the neighbourhood of $FeNi_3$ and the superlattice boundary suggested by them is incorrect. It should be extended for the majority of its length, to coincide with the binary Fe–Ni line.

(iii) *An Estimate of the Degree of Order*

It can be shown that the degree of order of an alloy may be calculated from a knowledge of the relative intensities of main and superlattice lines. The measurement of these relative intensities using the photographic technique was considered impracticable owing to the very great differences in intensities of the main and superlattice lines. In order to obtain a measurable intensity for a superlattice line the exposure required was so long that the blackening due to the main lattice lines was beyond the 'linear' range of the photographic emulsion.

The technique using a Geiger counter spectrometer as described, for example, by Hall, Arndt and Smith (1949) was considerably more sensitive than the photographic method. Consequently complete scans of the X-ray diffraction lines of two $FeNi_3$ specimens were made,* (*a*) an alloy quenched from 600°C and (*b*) an alloy cooled at 0·2°C per hour from 510°C. Cobalt $K\alpha$ radiation was used. No superlattice lines were observed with the quenched alloy. The areas of the diffraction lines were obtained by numerical methods using the experimental readings, suitably corrected for counting losses and secondary extinction. As the intensities of the superlattice lines obtained with the slowly cooled specimen were about 100 times less than those of the main lines, considerably longer counting times were employed for the superlattice lines. An accurate determination of the ratio of the intensities was, of course, impossible. However, from observations made it was concluded that the degree of order was in the range 55–80%. Without considerably improved X-ray technique this result represents the practically attainable limit of accuracy. The contrast between the line contours of the main and superlattice lines is clearly shown in fig. 6.

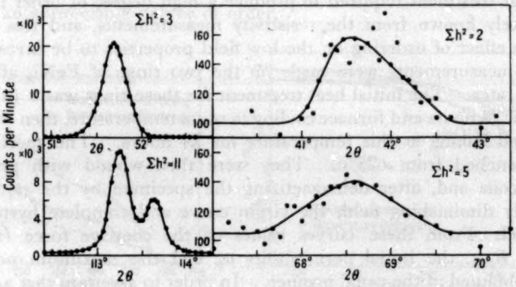


Fig. 6. Main and superlattice line shapes. The general background level for superlattice lines $\Sigma h^2 = 2$ and $\Sigma h^2 = 5$ was 105 ± 10 counts per min.

§ 5. MAGNETIC MEASUREMENTS

(i) *Effect of Ordering on Low Field Properties*

The magnetic properties of the iron–nickel alloys have been the subject of considerable investigation (see, for example, Bozorth 1951). The high permeability of alloys in the neighbourhood of 78% nickel, in low magnetic fields, is of considerable commercial importance. The existence of an order–disorder transformation at 75 atomic % nickel has led to a belief that this transformation is closely connected with the exceptional magnetic properties. Commercial high permeability alloys are too complex to enable a simple relation between the low field magnetic properties and the development of order to be readily obtained. Hence, in the investigations described in this paper, the low field properties of alloys of high purity were examined when in various degrees of order.

---

* The authors are indebted to Dr. R. A. T. Smith, Metallurgy Department, University of Birmingham, for kindly undertaking this section of the work.

Low field magnetic properties are notoriously inconsistent and difficult to reproduce; hence, to reduce the possibility of obtaining spurious results, two rings of pure $FeNi_3$ were used for all room temperature measurements. In order to eliminate demagnetizing effects, all specimens were machined to the anchor ring form. After forging to a solid cylinder and homogenizing, the centre portion of the ingot was parted from the outside on a lathe. A specimen for (σ, T) measurements was turned from the centre and a ring approximately 2·9 cm inside diameter, 3·6 cm outside diameter and 1 cm deep was turned from the outer portion. Ardron (1949) has investigated the effect of the dimensions of a ring on the low field magnetic properties and has found that if $t/D<\frac{1}{10}$, where t is the width of the metal of the ring and D the ring diameter, then the error in the measurement of these properties is less than 1%. All the rings used conformed to this condition, and in order to obtain a large cross-sectional area of metal the rings were made as deep as possible.

The conventional ballistic galvanometer method was used to measure the magnetic induction B produced by a field strength H in the specimen. Tests were made to ensure the absence of interference from the earth's and other stray magnetic fields.

The heat treatment required to produce a high degree of order in the alloy was accurately known from the resistivity measurements, and this knowledge enabled the effect of ordering on the low field properties to be investigated.

(B, H) measurements were made on the two rings of $FeNi_3$ after cooling at different rates. The initial heat treatment for these rings was: (*a*) annealing at 1250°c for 24 hours and furnace cooling to room temperature, then (*b*) reheating to 625°c and baking at this temperature for 24 hours. The specimens were first air-quenched from 625°c. They were then wound with primary and secondary coils and, after demagnetizing the specimen by the rapid reversal of a steadily diminishing field, the virgin curve and complete hysteresis loop were plotted. From these curves values of the coercive force H_c and the remanence B_{rem}, the initial permeability μ_0 and the maximum permeability $\mu_{\max}$ were obtained in the usual manner. In order to ascertain that any changes produced were solely due to the formation of order, the rings were heated rapidly to 520°c and held at this temperature for 24 hours before being furnace cooled. The effect on the low field properties of this treatment, which is known to produce no appreciable order in the alloy, and of cooling at rates which produce different degrees of order are shown in a typical set of results given in table 4.

Table 4. The Effect of Heat Treatment on the Low Field Magnetic Properties of $FeNi_3$

Heat treatment	(1)	(2)	(3)	(4)
μ_0	2270	1540	760	690
$\mu_{\max}$	8810	6460	4250	3820
H_c (oersted)	0·302	0·336	0·484	0·538
B_{rem} (gauss)	3570	3150	3150	3380

Heat treatments: (1) quenched in air from 625°c, (2) 24 hrs at 520°c and furnace cooled, (3) cooled at 1·37°c per hour from 510°c, (4) cooled at 0·29°c per hour from 510°c.

Examination of the results shows that both initial and maximum permeabilities are considerably decreased by heat treatments which are conducive to the formation of order. The values obtained for the remanence are not conclusive. The coercive force is definitely increased by the formation of order.

The values of the permeability of the slowly cooled alloys are in good agreement with those of Elmen (1928) but, owing to the shape of specimen used, his exact heat treatments could not be repeated in the present work, and the values of permeability for the rapidly cooled alloys reported here are much smaller than those of Elmen.

(ii) *The Effect of Ordering on the Coercive Force at High Temperatures*

Bennett (1949), in a study of the effect of ordering on the magnetic properties of Fe_3Al, found that, with low field measurements, the coercive force was the parameter most intimately connected with order–disorder changes. Measurements were therefore made on rings of $FeNi_3$ at elevated temperatures. High temperature insulation was the chief problem of this investigation. A ring of $FeNi_3$ was fitted with alumina tube insulators, one inside and one outside the ring, on which were wound primary and secondary coils of asbestos-covered wire. Insulation difficulties encountered at high temperatures were overcome by previously heating the asbestos windings in air to burn off carbon compounds in the insulation. The insulation was then very fragile and had to be handled with great care. The assembly was introduced into a large tubular vacuum furnace and cooled from 510°c to 400°c at 0·23°c per hour, a treatment known to produce an appreciable degree of order.

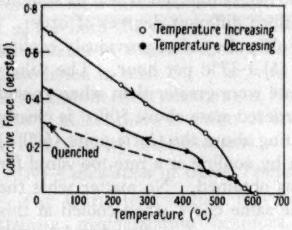


Fig. 7. Variation of coercive force with temperature, after cooling at 0·23° c per hour for $FeNi_3$.

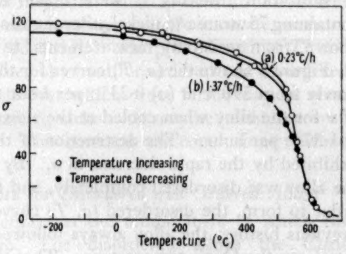


Fig. 8. Variation of saturation intensity of magnetization σ, with temperature for alloy containing 75 atomic % nickel which had been cooled at (*a*) 0·23°c per hour and (*b*) 1·37° c per hour.

The specimen was cooled to room temperature and measurements of the coercive force were made. Figure 7 shows the effect of heating the specimen up to the Curie temperature (600°c). Measurements were taken with the temperature increasing and decreasing, the temperature being held constant for about an hour before each reading in order that conditions should become steady. The curve for the ordered alloy may be seen to lie well above that for the disordered material. The fall in coercive force above 500°c, due to the destruction of the superlattice, is well marked, the value of H_c being zero at the

Curie point (600°c). To observe the effect of the formation of order on the coercive force, the ring was heated to about 490°c and maintained at this temperature for five days. The coercive force increased as expected. On quenching the same ring in water from above 600°c, a low value of the room temperature coercive force was obtained. On plotting this value in fig. 7 it appears that a certain amount of order was formed when the ring was cooled in the furnace.

(iii) *High Field Measurements*

The magnetic saturation intensity of a ferromagnetic alloy in the ordered state is different from that of the disordered alloy, and may be used, therefore, to examine the degree of order in such an alloy. Furthermore, the variation of the saturation intensity with temperature (the (σ, T) curve) will yield information on the state of order, will give a rough indication of the critical temperature, and will also enable the Curie temperature to be determined. Grabbe (1940) used single crystals of $FeNi_3$ to examine the effect of order on the magnetic anisotropy and saturation intensity at room temperature. Smoluchowski (1951) has recently discussed the increase in saturation intensity on the formation of order from the statistical viewpoint. Sucksmith (1951) reported on work done in his laboratory in which (σ, T) curves have been used to examine the degree of order in ferromagnetic alloys, viz. Fe_3Al, Fe_3Cr and $FeNi_3$.

Measurements of magnetic saturation intensity at high and low temperatures were made, using a ring balance of the type devised by Sucksmith (1939), who has described the principles of the method fully elsewhere. Measurements of saturation intensity σ were made at various temperatures on an alloy containing 75 atomic % nickel, after the alloy had been subjected to heat treatments known, from resistivity measurements, to produce different degrees of order.

Figure 8 shows the (σ, T) curves for this $FeNi_3$ alloy after previously cooling slowly from 510°c at (*a*) 0·23°c per hour and (*b*) 1·37°c per hour. The values of σ for the alloy when cooled at the slower rate were greater than when cooled at 1·37°c per hour. The destruction of the ordered state above 500°c is clearly exhibited by the rapid decrease in σ. By heating above the Curie point (600°c) the alloy was disordered completely, and then by cooling at a rate too rapid for order to form, the disordered (σ, T) curve was obtained. No matter what the previous history, the alloy always followed the same curve when cooled at this rate from the Curie temperature. That the alloy was in fact disordered by this treatment was confirmed by quenching a specimen in water from 700° c.

Specimens of iron–nickel alloys containing from 45 to 85 atomic % nickel were annealed for 24 hours at 600°c to remove the effects of cold work, and then cooled slowly from 510°c to 400°c at 0·23°c per hour to produce an ordered structure. Complete (σ, T) curves were obtained for each of the alloys by the foregoing procedure. As all the alloys had identical heat treatments, the results were directly comparable. Measurements were made on the ordered alloys at temperatures from room temperature to above the Curie temperature, and then down to room temperature, the cooling rate being too rapid for an appreciable degree of order to form. Figure 9 is shown as an example of the curves obtained, the alloys containing 70 and 81 atomic % nickel respectively.

The scatter of the individual points in all the (σ, T) curves was very small, indicating that the precision expected was attained.

It may be seen from figs. 8 and 9 that the difference in σ between the ordered and disordered alloys increases as the temperature increases to about 500° c, when disordering sets in. This is due solely to the fact that the ordered alloy has a higher Curie temperature than the disordered alloy and in no way implies that the degree of order is greater at these temperatures. When this difference in σ at various temperatures is plotted against the composition the maximum difference occurs close to 75 atomic % nickel, the difference being zero for all alloys at 600°c, indicating that disordering is complete at this temperature.

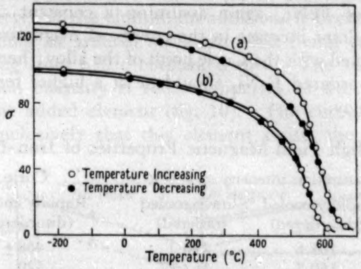


Fig. 9. Variation of saturation intensity of magnetization with temperature for alloys containing (*a*) 70 and (*b*) 81 atomic % nickel, after previous cooling at 0·23°c per hour.

The saturation intensity σ_0 at the absolute zero and the Curie temperature are inherent fundamental properties of magnetic materials and are structure-sensitive on a macroscopic scale. In order to obtain a more accurate measure of σ_0 for both the ordered and disordered states, σ was plotted against T^2 and extrapolated to absolute zero. The values obtained for ordered and disordered σ_0 are listed in table 5. The wide range over which order forms may be seen, the maximum effect occurring at 75 atomic % nickel. It may be noted that even at 50% nickel there is a sign of order. This confirms the conclusions reached in the resistivity measurements.

(iv) *Calculation of Curie Temperature for Disordered and Ordered Alloys*

The Curie temperature may be defined as the temperature at which the intrinsic magnetization becomes zero. For disordered alloys the Curie temperature was found by plotting σ^2 against T and extrapolating to $\sigma^2=0$. The extrapolation was short and the value of T for which σ^2 was zero gave the Curie temperature to within ±2°c.

The determination of the Curie temperature for the ordered state was not so simple since, with the iron–nickel alloys studied, disorder commences at about 500°c, one hundred degrees below the disordered Curie point. Extrapolation of the (σ, T) curve was extremely inaccurate. In order to make the ordered and disordered curves more comparable a graph of the reduced magnetization σ/σ_0 was plotted against the absolute temperature. The curve for the disordered alloy was plotted to the Curie point, whereas the ordered points could only be plotted to about 500°c. The extrapolation to $\sigma/\sigma_0=0$ was still rather long and hence subject to considerable error. The method finally adopted was to plot $(\sigma/\sigma_0)^2$ against T^2, the extrapolation in this case being much smaller, giving the ordered Curie temperature to within ±5°c.

The Curie temperatures obtained in this manner are listed in table 5. When plotted, the curve for the disordered alloy is in good agreement with previously published work. There was, as expected, more scatter among the ordered points. However, the graph shows a definite maximum, indicating that the magnetic energy is considerably increased on the formation of the superlattice. The increase in the Curie temperature with ordering is to be expected on theoretical grounds. Bethe (1933) and Slater (1930) have related theoretically the energy of magnetization to the interatomic distance, assuming a constant diameter for the 3d shell in all ferromagnetics. A decrease in the atomic separation of $FeNi_3$, again assuming a constant diameter for the 3d shell, implies a resultant increase in the energy of magnetization. This energy is intimately connected with the Curie point of the alloy; hence by this reasoning the Curie point of ordered $FeNi_3$ should be at a higher temperature than that of the disordered alloy.

Table 5. High Field Magnetic Properties of Iron–Nickel Alloys

	Saturation intensity σ_0 at 0°K		Curie temperature (°C)	
Alloy composition (atomic % Ni)	Rapidly cooled (disordered)	Slowly cooled (ordered)	Rapidly cooled (disordered)	Slowly cooled (ordered)
45	168·6	170·3	468 ± 2	494 ± 5
50	160·5	162·1	520	543
55	152·4	154·0	558	580
60	143·6	144·7	592	616
65	133·4	135·9	613	636
68	128·6	131·7	616	668
70	124·0	128·0	614	680
72	120·0	124·6	608	696
74	115·8	121·0	600	691
75	113·4	118·8	598	681
76	111·3	116·0	589	654
78	106·4	107·7	585	624
80	101·9	102·9	577	599
81	98·8	100·6	571	584
85	90·2	91·0	543	543

§ 6. EFFECT OF THE ADDITION OF FOREIGN ATOMS

(i) *Experimental*

The addition of a small percentage of a third element to iron–nickel alloys is known to influence considerably the physical properties of the alloy. The maximum effect of the order–disorder transformation occurs at $FeNi_3$, hence this composition was used as the base to which small percentages of other metals were added, i.e. the ratio of the number of nickel atoms to the number of iron atoms in the ternary alloy was maintained at 3 : 1. The amount of the third element added was calculated in atomic %, to simplify any theoretical implications. The measurement of (σ, T) curves proved to be the most informative and also the most rapid method of examining the order–disorder phenomenon, and so was used to study the ternary alloys.

The addition of up to 6 atomic % of aluminium has been shown (by x-ray technique) to assist the formation of the superlattice (Bradley and Taylor 1940). Molybdenum and copper have been added extensively in industry to iron–nickel alloys to form the molybdenum–permalloys, Mu-metal, etc. Manganese is often added to commercial alloys as a deoxidizer and is of considerable scientific interest as it is the element which is responsible for the ferromagnetism of the Heusler alloys.

After removal of the effects of cold work (by annealing for 24 hours at 650°c) the alloys were obtained in the ordered state by cooling slowly from 510°c to 400°c at 0·23°c per hour. The saturation intensity was measured at temperatures from room temperature to 650°c, a temperature at which the alloys were fully disordered. Measurements were then made as the temperature was reduced to room temperature and the (σ, T) curve plotted for each alloy. From the (σ, T) curves of the aluminium additions it became apparent that the 6% alloy was forming an ordered structure while being cooled in the balance furnace ; quenching from 650°c confirmed this. The values of ordered and disordered saturation intensity at room temperatures were plotted against the concentration of the added element (fig. 10). The curves for the aluminium additions show conclusively that this element assists the formation of order. Molybdenum and copper affect neither the rate of formation of order, as the values for the quenched alloys coincide with those for the alloys when cooled in the balance furnace, nor do they affect the degree of order obtained by a given heat treatment, since the difference in σ between the ordered and disordered states is the same for all compositions. Molybdenum and copper both decreased the saturation intensity in the ordered and disordered states in a regular manner. With manganese, however, the effect was more complex. For the disordered alloy the addition of manganese had practically no effect on the value of σ at room temperature, whereas, for the ordered alloy, σ was increased.

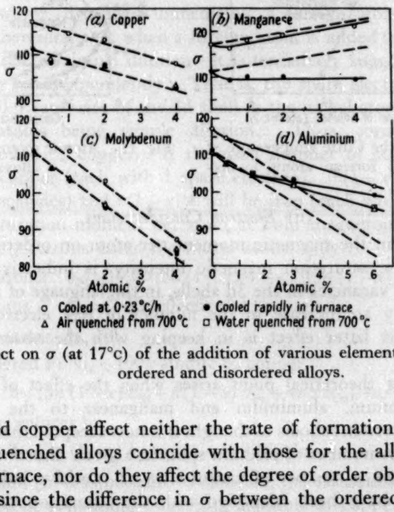


Fig. 10. Effect on σ (at 17°c) of the addition of various elements to $FeNi_3$ for ordered and disordered alloys.

The values of the disordered Curie temperatures were calculated from the (σ, T) curves and are plotted against the concentration of foreign atoms in fig. 11, together with the effect of adding iron and nickel to $FeNi_3$.

The outstanding feature of this graph is that the addition of copper has practically no effect on the Curie temperature of $FeNi_3$. Molybdenum and aluminium decrease the Curie temperature, molybdenum more than aluminium.

The aluminium alloy containing 6 atomic % aluminium was impossible to obtain in the disordered state by cooling in the furnace, hence the value obtained was greater than if it were completely disordered. The behaviour of manganese is anomalous in that the addition of up to ½ atomic % increases the Curie temperature, whereas for additions greater than this the Curie temperature is decreased.

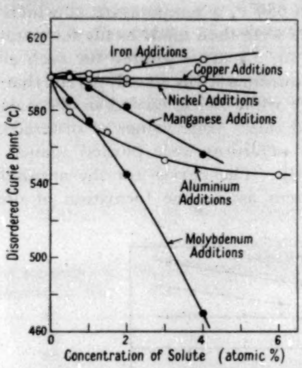


Fig. 11. Effect on the Curie temperature of the addition of foreign atoms to disordered $FeNi_3$.

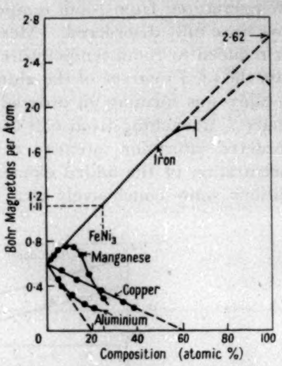


Fig. 12. Atomic moments of some nickel alloys.

(ii) *Electron Considerations*

The increase in the magnetic moment per atom on ordering, as shown by an increase in the saturation magnetic intensity, is indicative of an increase in the number of vacancies in the 3d shells, in the language of atomic structure, and thus possibly an increase in the number of free electrons available for conduction. This latter effect is in keeping with the observed decrease in resistivity.

An interesting theoretical point arises when the effect of the addition of copper, molybdenum, aluminium and manganese to the alloy $FeNi_3$ is considered. In certain schools of thought it is now accepted that ferromagnetism is due to an excess number of electron spins in a given direction in the 3d shell. There is, therefore, a number of vacant sites or 'positive holes' in this shell. These vacancies have been identified with the saturation magnetic moment in Bohr magnetons, i.e. 0·6 per atom for nickel.

Stoner (1933) and Mott (1935) have shown a strikingly simple correlation between experiment and theory when copper is added to nickel. Copper–nickel alloys form face-centred cubic structures for all compositions. The copper atom (atomic number 29) contains one more electron per atom than nickel (atomic number 28). Therefore, to a first approximation, if a nickel atom is replaced by a copper atom the lattice may be supposed to remain unaltered except that an extra electron is added. If it is assumed that these extra electrons enter the 3d nickel shell, the saturation moment per atom should be decreased until, with an alloy containing 60% copper, the 3d shell is full; then the saturation moment should be zero. Figure 12 shows that this is confirmed experimentally.

The effect on the magnetic moment of the addition of other metals to nickel is included in the figure.

It may be seen from the figure that, for alloys in the region $FeNi_3$, the contribution of iron to the atomic moment may be taken as 2·62 Bohr magnetons (the atomic moment of pure iron is 2·2).

If we assume that the iron and nickel atoms act as a single unit, then the atomic moment of $FeNi_3$ may be calculated from the moments of iron (2·62) and of nickel (0·6) thus: magnetic moment of $FeNi_3 = b = \frac{1}{4}\,[3 \times 0{\cdot}6 + 2{\cdot}62] = 1{\cdot}11$. This result is confirmed by the value of σ_0 determined by the experiments described earlier in this paper.

Now, treating $FeNi_3$ as a single unit and considering the addition of copper in the same way as in the case of the addition of copper to nickel referred to above, the expected reduction in the saturation intensity does not agree with that obtained experimentally.

The work of Sadron (1932 a, 1932 b) and Fallot (1938) on the addition of various elements to iron shows that the effect of such additions is one of pure dilution for small percentages of the solute. The magnetic moment of the nickel atom, however, decreases immediately a foreign atom is added to it. Having shown, therefore, that when a foreign atom is added to the alloy $FeNi_3$, the effect is not one of simple dilution—it is tentatively suggested that the iron and nickel atoms act independently. That is, the spare electrons of the added atoms enter the 0·6 vacancies of the 3d shell in the nickel atoms only, the effect with the iron atoms being simple dilution. Thus, consider an alloy of $FeNi_3 + 100\,x$ atomic % copper. If the total number of atoms present is n, then nx will be copper (each with 1 spare electron), $3n(1-x)/4$ will be nickel (each with 0·6 vacancies) and $n(1-x)/4$ will be iron (each with 2·62 vacancies).

If b is the saturation moment per atom in Bohr magnetons, then, assuming the spare copper electrons enter the nickel atoms only :

for the nickel atoms $b_{Ni} = [0{\cdot}6 \times 3n\,(1-x)/4 - n\,x]/3n\,(1-x)/4$, ... (1)

for the iron atoms $b_{Fe} = 2{\cdot}62$ (effectively),

for the copper atoms $b_{Cu} = 0$.

Hence for disordered $FeNi_3 + 100x$ atomic % copper

$$b = [3n\,(1-x)\,b_{Ni} + n\,(1-x)\,b_{Fe} + 4n\,x\,b_{Cu}]/4n,$$

which in our case reduces to

$$b = [3n\,(1-x)\,b_{Ni} + n\,(1-x)\,2{\cdot}62]/4n. \qquad \ldots\ldots(2)$$

From eqns. (1) and (2) it is possible to calculate the saturation intensity for the alloys of pure $FeNi_3$ containing increasing amounts of copper.

In the case of pure $FeNi_3$ the saturation intensity at room temperature is increased by a factor 1·055 on ordering by cooling at 0·23°C per hour (see fig. 8). Hence, to a first approximation, the theoretical values of b' for the ordered ternary alloys may be obtained from the corresponding values of b for the disordered alloys by multiplying by this factor.

The theoretical curves for ordered and disordered material, drawn in fig. 10 (*a*), together with the experimental points, indicate that the assumptions made were fully justified. Theoretically σ_0, the value at absolute zero, should have been used for this calculation, but the room temperature value differs only slightly from σ_0 and can be more accurately measured.

It has been suggested that the six outermost electrons of molybdenum enter the 3d sub-shell of nickel when molybdenum is added to pure nickel (Wohlfarth 1949). Thus molybdenum is six times more effective than copper in reducing the saturation intensity of nickel. By a similar calculation good agreement with experiment was again obtained (fig. 10 (*c*)).

The aluminium atom has three spare electrons per atom. The theoretical curves for these additions (fig. 10 (*d*)) agree with experiment only for very dilute solutions of aluminium. With the addition of aluminium to pure nickel (fig. 12) the experimental curve deviates from the theoretical in a similar manner. This, it is suggested, is due to the formation of an ordered structure of the type Ni_3Al as, from the work described in this paper, the formation of order increases σ.

The behaviour of manganese is not so simple. Néel (1940) has shown that the algebraic increase in the number of Bohr magnetons, when one atom of nickel is replaced by one atom of manganese, is $+2{\cdot}4$ for small concentrations of the solute. Thus the addition of manganese to $FeNi_3$ should theoretically increase σ as shown in fig. 10 (*b*). The agreement with the ordered alloys is reasonable, but for the disordered alloys the saturation intensity is unchanged by the addition of manganese in contradiction to theory. The strange behaviour of manganese, which has such a complex structure, has caused considerable interest in the past (e.g. Guillaud 1944).

§ 7. GENERAL CONCLUSIONS

From the evidence presented there can now be little doubt that an ordered state exists for a wide range of iron–nickel alloys. It has been shown that, from measurements of the resistivity, lattice parameter, saturation intensity of magnetization and Curie temperature, on alloys ranging from 45–85 atomic % nickel the largest difference in the value of these properties for different heat treatments occurs at 75 atomic % nickel. The occurrence of this maximum effect in such simple stoichiometric proportions is in agreement with the atomic ratio at which other superlattices are found.

The theory of order–disorder phenomena (Bragg and Williams 1934, 1935 a, 1935 b, Nix and Shockley 1938, etc.) for face-centred cubic superlattices with a 3 : 1 atomic ratio has been based mainly on experimental data obtained from a large number of investigations of alloys of the type $AuCu_3$ in which a superlattice readily forms. For example, theory predicts, for an alloy of this type, a sudden decrease in the resistivity just below the critical temperature, a phenomenon well exhibited by $AuCu_3$. For $FeNi_3$, however, a similar face-centred cubic superlattice is produced, but such a sudden decrease below the critical temperature is not observed. This, it may be argued, is not necessarily indicative of the theory being inadequate, but rather that, the transformation being extremely sluggish, the times necessary to note this decrease become impractically long.

It is surprising that the range of composition over which the order–disorder transformation appears to take place is so wide. Definite evidence has been obtained that the ordered structure extends over a range from 65 to 81 atomic %, with the tentative suggestion that it may extend even to compositions as low as 45% nickel. As far as is known at present no alloy system has such a wide range of composition, presumably with the same structure, over which order may exist. If there is indeed an ordered structure at 50% nickel it is difficult to see how it can still take the face-centred cubic form, but no evidence has yet been advanced of any other structure at that composition in the single phase. The change in lattice parameter with composition does not indicate a mixed phase region. On the other hand, Owen and Liu (1949) have produced evidence to support the claim that the mixed ($\alpha+\gamma$) phase extends to about 60 atomic % nickel. It has been suggested in § 3 that the resistivity curve indicated the possibility of the development of another phase. If this is so it would account for the decrease in the lattice parameter on prolonged annealing, but it would not explain the decrease over the range of composition obtained in this work. An attempt is being made to investigate this region further.

Hoselitz (1944) has produced evidence to suggest an ordered structure at 25 atomic % nickel, but any confirmatory investigation must be difficult because of the presence of the α-phase.

The x-ray (or neutron) diffraction technique is the only positive method of proving the existence of a superlattice. In the case of the iron–nickel alloys, however, the method is considerably handicapped by the similarity in scattering factors of the constituent atoms. This is true also of neutron beam scattering. The change in cell dimensions resulting from the development of order is small, and thus the x-ray technique is severely limited when applied to an examination of the order–disorder transformation in these alloys. The measurement of electrical resistivity and the examination of the magnetic properties are more satisfactory parameters with which to follow the transformation.

The initial permeability μ_0 and coercive force H_c are structure sensitive properties of magnetic materials, i.e. they may change in value by large amounts as a result of small changes in the heat treatment. On the basis of the modern theory of magnetic domains an increase in the internal stress of a metal causes an increase in H_c and a decrease in μ_0. One would expect that, on the formation of order, internal stresses on an atomic scale would decrease, which suggests that H_c would also decrease on ordering. This, however, is contrary to experimental evidence. Sucksmith has suggested that the reason for this may lie in the formation of antiphase nuclei, i.e. small islands of ordered material in a matrix of disordered material. The boundaries between ordered and disordered material will be regions of relatively high energy and sources of internal strain. Experimental evidence of these antiphase nuclei has been suggested by the diffuse nature of the x-ray diffraction lines of the superlattice, even after extremely slow cooling.

It has been shown that the order–disorder transformation does not bring about an increase in the permeability of iron–nickel alloys, and it is unlikely that any appreciable degree of order will be formed in commercial alloys when given the usual heat treatments, as has frequently been suggested in the literature.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to Messrs. Wm. Jessop Ltd., the Mond Nickel Co. Ltd., and the Telegraph Construction and Maintenance Co. Ltd. for supplying them with alloys of commercial purity on which the initial measurements were made; also to Prof. W. Sucksmith for helpful suggestions and for providing ample facilities in the Physics Department where this work was carried out. One of us (R. J. W.) also wishes to thank Henry Ellison, Esq., for providing financial aid in the form of an Ellison Fellowship, without which it would not have been possible to undertake this work.

References

Ardron, G. H. C., 1949, *Thesis*, University of Sheffield.

Bennett, W. D., 1949, *Thesis*, University of Sheffield.

Bethe, H., 1933, *Handbuch der Physik*, **24,** 595.

Bozorth, R. M., 1951, *Ferromagnetism* (New York : D. van Nostrand), Chapter 5.

Bradley, A. J., Jay, A. H., and Taylor, A., 1937, *Phil. Mag.*, **23,** 545.

Bradley, A. J., and Taylor, A., 1938, *Proc. Roy. Soc.*, **166,** 353; 1940, *J. Inst. Met.*, **66,** 53.

Bragg, W. L., and Williams, E. J., 1934, *Proc. Roy. Soc.*, **145,** 699; 1935 a *Ibid.*, **151,** 540; 1935 b, *Ibid.*, **152,** 231.

Dahl, O., 1936, *Z. Metallkde.*, **28,** 133.

Easthope, E. E., 1937, *Proc. Camb. Phil. Soc.*, **33,** 502.

Elmen, G. W., 1928, *J. Franklin Inst.*, **206,** 317.

Fallot, M., 1938, *Ann. Phys., Paris*, **10,** 291.

Grabbe, E. M., 1940, *Phys. Rev.*, **57,** 728.

Guillaud, C., 1944, *C.R. Acad. Sci., Paris*, **219,** 614.

Hall, W. H., Arndt, U. W., and Smith, R. A., 1949, *Proc. Phys. Soc.* **A, 62,** 631.

Hoselitz, K., 1944, *J. Iron Steel Inst.*, **149,** 193.

Källbäck, O., 1947, *Ark. Mat. Astr. Fys.*, **34B,** No. 17.

Kaya, S., 1938, *J. Fac. Sci. Hokkaido Univ.*, Ser. II, Bd. **2,** 29.

Leech, P., and Sykes, C., 1939, *Phil. Mag.*, **27,** 742.

Marsh, J. S., 1938, *Alloys of Iron and Nickel* (New York : McGraw-Hill), Vol. 1, p. 342.

Mott, N. F., 1935, *Proc. Phys. Soc.*, **47,** 571.

Néel, L., 1940, *Le Magnétisme*, **11,** 65.

Nix, F. C., and Shockley, W., 1938, *Rev. Mod. Phys.*, **10,** 1.

Owen, E. A., and Liu, Y. H., 1949, *J. Iron Steel Inst.*, **163,** No. 3, 132.

Owen, E. A., Sully, A. H., and Yates, E. L., 1937, *Proc. Phys. Soc.*, **49,** 318.

Sadron, C., 1932 a, *Thesis*, University of Strasbourg; 1932 b, *Ann. Phys., Paris*, **17,** 371.

Slater, J. C., 1930, *Phys. Rev.*, **36,** 57.

Smoluchowski, R., 1951, *Phys. Rev.*, **84,** 513.

Stoner, E. C., 1933, *Phil. Mag.*, **15,** 1018.

Sucksmith, W., 1939, *Proc. Roy. Soc.*, **170,** 551; 1951, *J. Phys. Radium*, **12,** 430.

Wohlfarth, E. P., 1949, *Phil. Mag.*, **40,** 1109.

Yates, E. L., 1946, *J. Sci. Instrum.*, **23,** 229.

The Freezing of Supercooled Water in Glass

By R. G. WYLIE
Commonwealth Scientific and Industrial Research Organization, National Standards Laboratory, Chippendale, N.S.W., Australia

Communicated by F. C. Frank; MS. received 1st July 1952, and in amended form 17th October 1952

Abstract. The paper gives an account of some experiments in which water was supercooled and frozen in glass. Elaborate precautions, which were taken to eliminate possible effects due to the colloidal matter of the atmosphere, permitted the study of freezing in an unusually simple environment. Comparison with the behaviour of water frozen in contact with the atmosphere allows some assessment of the significance of airborne contamination. The results, including those of some incidental experiments on the nucleating abilities of graphite, dust and the abrasion of glass on glass, are considered in relation to theory. It is found that the use of glass for the retaining surface precludes the study of spontaneous nucleation in the bulk of the water.

---

§ 1. INTRODUCTION

In the great majority of cases, when the independent variables of a thermodynamic system are varied so that a boundary line in the phase diagram is crossed, no new phase at first appears but a metastable state results. The metastable state can be destroyed by certain nucleation processes which, except perhaps in the case of condensation of a pure vapour, are not well understood. The supercooling of water provides an important example which has largely resisted the efforts of both theoretical and experimental investigators for more than half a century. Reliable data concerning the nucleation of ice in supercooled water may be expected to have considerable significance for meteorology as well as for our knowledge of the structure of matter which must be intimately involved in the theoretical interpretation.

§ 2. PREVIOUS EXPERIMENTAL WORK

The literature on the properties of supercooled water is extensive, having expanded considerably both during and since the war. However, the various conclusions which have been put forward are by no means unequivocal and do not admit of being paraphrased in a few words. For present purposes attention may be confined to experiments in which glass vessels were used, except to remark that the threshold temperature below which self-nucleation in the body of the liquid would become easily observable can hardly lie above −40°c (Sander and Damkohler 1943, Fournier d'Albe 1948). An extensive bibliography has been given by Dorsey (1948).

The freezing of supercooled water in glass has been ably investigated by Tammann and Buchner (1935) and Meyer and Pfaff (1935). Tammann and Buchner found that freezing temperatures were frequently characteristic of the specimen and that freezing often began from the same region of the tube. From the different behaviour of different specimens they inferred the presence of foreign nucleating matter.

Meyer and Pfaff focused their attention on the freezing nuclei. They found that considerably lower temperatures could be produced by filtering, the lowest obtained being −33°C. By repeated vacuum distillation without ebullition, in glass apparatus which had been 'steamed out and carefully cleaned', they obtained samples of successively higher purity which froze in different ways. They attributed the freezing in their best samples to rough crystalline places on the glass surface. Among numerous other investigations the extensive work of Dorsey (1948) may be mentioned.

The results obtained by these workers include no quantitative data of general significance, which epitomizes the character of the literature generally. The outstanding need is for reproducible data, even if these are fundamentally statistical or relate to secondary effects. Record low freezing temperatures are interesting but of little value unless reproducible.

With the exception of various expansion chamber experiments not explicitly mentioned above, all the freezing experiments hitherto reported have allowed contamination by the suspended matter of the atmosphere. This is less obviously true of experiments such as those of Meyer and Pfaff, who made use of repeated vacuum distillation; but, however well purified the water may be, if the wall of the vessel in which it is contained has been exposed to the atmosphere then the contamination is present. As is easily demonstrated with a simple expansion chamber, the atmosphere commonly contains thousands of condensation nuclei per cubic centimetre, and under laboratory conditions sometimes much larger numbers. The nature of these, and their number, depends on the environment. Even allowing that many of them are soluble, the possibility cannot be ignored that a considerable number may be capable of nucleating the freezing process, at least at some temperature before self-nucleation sets in. Chemical cleaning of the apparatus following conventional technique will not suffice, for, as the chemicals are poured out, the atmosphere again enters, and its entrained particles diffuse to the walls, where they adhere. The principal object of the experiments described below was to determine whether this atmospheric contamination was the limiting factor in supercooling water in glass.

§ 3. THEORETICAL

The formation of crystal embryos in a supercooled melt is essentially a temperature activated rate process. In the formation of a crystal by the chance aggregation of molecules the free energy of the system passes through a maximum. If the liquid remains homogeneous this corresponds to that size of crystal embryo which is in unstable equilibrium with the melt. If the equilibrium embryo is treated as a sphere, then, assuming that the molecular Helmholtz free energy and the molecular volume are the same inside the embryo as in a macroscopic crystal and with no further assumption, it is easily shown, by minimizing the thermodynamic potential, that the condition for equilibrium is

$$\psi_\beta - \psi_\alpha = \frac{2\gamma v_\alpha}{r} + P(v_\alpha - v_\beta) \qquad \ldots\ldots(1)$$

where ψ_β and ψ_α are respectively the molecular Helmholtz free energies of water and ice, v_β and v_α are the respective molecular volumes, r is the radius of the embryo, γ is the interfacial free energy and P is the pressure in the liquid.

The term in (1) which represents the work done by the pressure is quite negligible for ordinary pressures. When this is dropped and the difference of the free energies is replaced by its expansion in terms of δT at constant pressure, δT being the deviation of the temperature from the macroscopic melting point, the result is

$$-\frac{L_\mathrm{m}}{T_\mathrm{m}}\delta T-\tfrac{1}{2}\frac{(C_{\beta,\mathrm{m}}-C_{\alpha,\mathrm{m}})}{T_\mathrm{m}}(\delta T)^2=\frac{2\gamma v_\alpha}{r}, \qquad \ldots\ldots(2)$$

where L_m is the latent heat, $C_{\beta,\mathrm{m}}$ and $C_{\alpha,\mathrm{m}}$ are respectively the specific heats of water and ice and where the subscript m denotes the melting point. This equation differs from the well-known result of J. J. Thomson (1888) in possessing the term in $(\delta T)^2$. For water the latter term is 13% of the first-order term when $\delta T=-40$°c. A close numerical agreement between experimental results and Thomson's equation, as obtained by Lafargue (1950), thus appears to be fortuitous, the effect of omitting the second-order term presumably being cancelled by errors in the estimated values of γ and r.

The remarkable agreement between the Becker–Döring theory of vapour nucleation (1935) and the experimental results of Volmer and Flood (1934) has shown that it is not seriously incorrect to use the macroscopic surface tension of water for a droplet containing only about 80 molecules. However, in the theory of freezing somewhat smaller aggregates are in question and some reservation must be made on this point. Nevertheless the theory must not be discarded simply on account of this doubt, which has been greatly laboured in the literature.

No complete kinetic calculation has been given for the crystallization process on the basis of any model. A theory has been given by Frenkel (1932) which is based on the quasi-thermodynamic treatment of vapour condensation due to Volmer and Weber (1926). The latter workers argued, on the basis of Einstein's fluctuation theory (1910), that the probability per cm^3 per second that a droplet will form in a supersaturated vapour is given by

$$J=c\exp(-\Delta F_0/\boldsymbol{k}T) \qquad \ldots\ldots(3)$$

where ΔF_0 is the work of formation of the droplet of equilibrium size, $\boldsymbol{k}$ is Boltzmann's constant, T is the absolute temperature and c is a constant. The use of this equation for the crystallization process is vitiated by the existence of considerable viscous resistance to the free supply of molecules at the surface of the embryo, a complication which presents a formidable obstacle to the production of a true kinetic theory. Frenkel, noting that the pre-exponential constant in (3) contained as a factor the number of molecules available at the interface for addition to the crystal, suggested that the viscosity effect was taken into account if only molecules in an activated state were counted. He accordingly replaced C by $C'\exp(-U/\boldsymbol{k}T)$, where U is the activation energy for viscous flow and C' is a new constant. His theory is thus summarized by the equation

$$J=C'\exp\{-(U+4\pi r^2\gamma/3)\} \qquad \ldots\ldots(4)$$

where ΔF_0 has been replaced by the value $4\pi r^2\gamma/3$ which, by means of a very general argument, it is easily shown to possess.

The interfacial free energy γ is beyond present experimental techniques to measure, and for present purposes may be estimated following Volmer's suggestion (1939, p. 181), the assumption being that it bears the same ratio to the surface tension of water as the latent heat of fusion does to the latent heat of vaporization of water. A value of 10·5 erg/cm^2 is obtained in this way.

A curve of log J/C' plotted against δT for water, with $\gamma = 10$, is given in fig. 1. A theory put forward by Turnbull and Fisher (1949), which is based on the usual method of calculating absolute reaction rates, gives for C' a value of roughly 10^{36}. With the value of γ adopted here, this figure would preclude the possibility of supercooling water more than a few degrees, so that either the factor obtained from their theory is much too large or the adopted value of γ is substantially too low. The other curves in the figure arise from the following considerations.

The nucleation of a supercooled liquid by foreign crystals is highly specific; ice nucleates water at a temperature very close to 0°c but no other substance is known which causes nucleation at a temperature higher than −5°c. Silver iodide, which is pseudomorphic with ice, is effective for temperatures below about −7°c. It is not *a priori* obvious that an ice crystal should nucleate supercooled water easily; indeed if the crystal were perfect it almost certainly would not (Frank 1949). Thus imperfection of the crystal surface is involved. It is very reasonable to expect that when a foreign crystal, especially a pseudomorphic crystal, is introduced into a supercooled liquid and causes nucleation, its surface imperfections play an analogous role. Thinking in terms of a spherical model, it is easy to see that whereas ice must 'wet' ice, the specificity of nucleation catalysis implies that ice does not 'wet' foreign crystals, but rather that the angle of contact is in general large. It is easy to show that when the contact angle is $\frac{1}{2}\pi$ the free energies of formation of equilibrium embryos on a perfect plane at the rectangular intersection of two planes and in a rectangular corner are $\frac{1}{2}\Delta F_0$, $\frac{1}{4}\Delta F_0$ and $\frac{1}{8}\Delta F_0$ respectively. Corresponding curves of log J/C' have been given in fig. 1. These sites represent typical defects of a crystal surface. When the equivalent contact angle is less than 54° 48′ nucleation in a rectangular corner requires no supercooling unless a linear free energy of the boundary line is invoked. Likewise at the rectangular intersection of two planes no supercooling is required when the equivalent contact angle is less than $\frac{1}{4}\pi$.

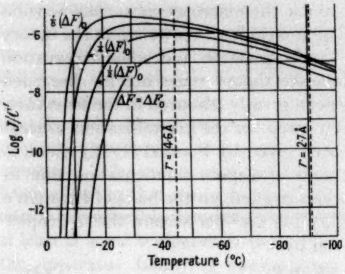


Fig. 1.

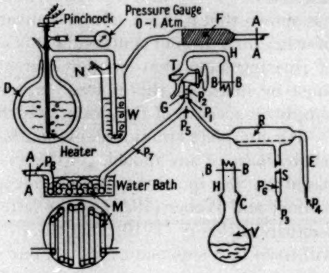


Fig. 2.

§4. Experimental

The principal piece of apparatus used is shown diagrammatically in fig. 2. An oxygen stream, which was provided by evaporation of liquid oxygen from the Dewar vessel D, and which passed through triply distilled water W and then through a long sintered-glass filter F, flowed through the apparatus throughout an experiment. This gas stream was tested for condensation nuclei using a small expansion chamber, and found to contain none.

After being filtered, the gas passed over triply distilled water in the glass unit M, which was termed the 'maze'. No bubbling or gas bubble formation was allowed to occur in this. At the appropriate time the temperature of the maze could be raised slowly to 70–80°c to increase the water-vapour content of the gas stream so that liquid water could be condensed out from the gas in the reservoir R, which was temporarily provided with a cold jacket for the purpose. The glass tube bearing the oxygen and the reservoir had three appendages. The first consisted of the tube G, the two-way tap T, which was lubricated only with distilled water, and the flask V, and was associated with acid and steam cleaning of the reservoir and of the tubes S and E. The latter tubes, which were the other two appendages, provided either alternative or simultaneous escapes to the outside atmosphere for the oxygen stream, and tube S was the actual sample tube in which water was condensed and cooled. The cleaning operations, which involved raising the sample tube to a temperature above its softening point while the oxygen flowed through it, were carried out as follows.

Beginning with the glass tube G terminating in an open end at P_1 and with the branch S sealed off at P_3, the reservoir was filled through G to the level of the dotted line with the desired cleaning acid; both pure nitric acid and chromic acid were tried. The glass connection was made at P_1. The oxygen then being allowed to flow, the glass between P_1 and P_2 was strongly fired while the gas was temporarily allowed to discharge through both H and E and the tap T was then closed. The flask, containing doubly distilled water, was fitted at the cone-and-socket joint C which, with the tap T, was sealed with distilled water. The acid was then heated with a hand torch, objectionable fumes being removed at P_4 by means of a water-pump aspirator. The oxygen stream prevented any acid vapour passing back to or beyond the point P_5. The water in the flask V having meanwhile been brought to the boiling point, the tube H was flushed out to the atmosphere with steam through the tap T and the stream then turned into the reservoir. A strong flow of steam was allowed to develop. The seal at P_3 was broken with the aid of a glass knife and forceps and the hot acid allowed to discharge from a short distance into a large volume of cold water. The steam was then allowed to issue for about half an hour from both (similar) jets P_3 and P_4. At the end of this time it was diverted to the atmosphere by tap T. For convenience the glass tube G was usually drawn off between P_1 and P_2, using no more heat than necessary. The oxygen was then discharging quietly from the jets P_3 and P_4.

The next stage in the cleaning was to heat the glass tube to a temperature above its softening temperature, either with a small tubular furnace or with a hand torch. The tubular furnace was used when the tube S was straight as shown in fig. 2 and hand-torching when the S-shaped tube shown in fig. 3 was used.

When the straight tube was used with the concentric tubular furnace, the gas issuing from P_3 was continuously sampled with a specially developed technique and tested with a small expansion chamber. Until heating was begun, the issuing gas always remained entirely free of condensation nuclei. Shortly after the commencement of heating, large numbers of nuclei were observed to form water drops in the expansion chamber. As the temperature reached about 400°c the number of nuclei had become so great that the fogs produced were much finer than the finest which are capable of giving coloured rings of the type described by Wilson (1897).

At 600°c the number of nuclei was decreasing very rapidly, and at 700°c was zero. Indeed, when the temperature had been held at 700°c for about half an hour not a single condensation nucleus could be detected in the issuing oxygen stream. Also it was found that if heating were carried to, say, 600°c and many nuclei were still being obtained, and if the temperature were then immediately lowered to 550°c, the number of nuclei in the oxygen dropped practically to zero. Further, once the tube had been maintained at 700°c for some minutes any further increase in temperature up to 770°c gave rise to very few condensation nuclei. At higher temperatures than 770°c the tube elongated appreciably under its own weight.

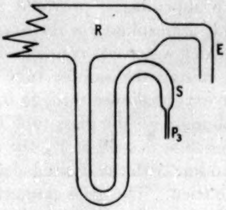


Fig. 3.

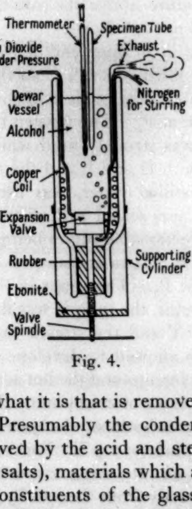


Fig. 4.

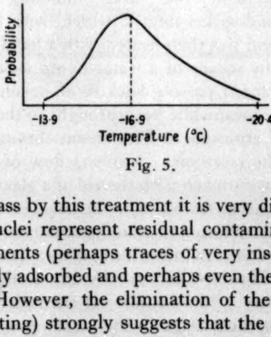


Fig. 5.

Just what it is that is removed from glass by this treatment it is very difficult to say. Presumably the condensation nuclei represent residual contamination not removed by the acid and steam treatments (perhaps traces of very insoluble oxides or salts), materials which are strongly adsorbed and perhaps even the more volatile constituents of the glass itself. However, the elimination of the effect at roughly 700°c (for several minutes heating) strongly suggests that the nuclei do arise from surface matter and not from the glass, and that this matter is properly removed, probably both by evaporation and by solution in the glass. If the contamination were regarded as having a whole 'continuous spectrum' of vapour pressures then the solution theory would be very probable since the elimination of the effect coincides with the onset of obvious plasticity of the glass. The tube S was heated until condensation nuclei no longer occurred for each of three overlapping positions of the furnace. Only the length of glass covered in the central position was utilized in the preparation of samples.

When the S-shaped tube was employed it was heated by hand-torching to as high a temperature as practicable, and with the above results in mind. In this case a still further cleaning operation was carried out. The tip at P_3 was sealed by touching it with an oxy-gas flame (the glass would seal completely against the oxygen-flow pressure by surface tension contraction) and water was then condensed quietly into the U-section of the S-tube by raising the temperature of the maze and cooling the reservoir, the oxygen issuing at P_4. Then, on breaking the tip at P_3 with forceps, as before, the water was blown out by the gas; the S-tube could be washed out with extremely clean water several times in this way. The cleaning was then regarded as completed.

The samples to be cooled were condensed into the straight tube after it had been sealed off at the point P_6 or into the U-section of the S-tube after its tip P_3 had been sealed by surface tension contraction as mentioned above. The condensation process was observable only by the gradual accumulation of water in the sample tube, no bubbles forming in the maze and actual condensation occurring in the reservoir in an invisible surface film.

In the earlier experiments an additional branch tube joining the main tube at P_7 (not shown in fig. 1) was connected through a glass tap and a liquid-air trap to a rotary vacuum pump. The tap was lubricated with Apiezon grease L. Other differences of a purely geometrical character existed. The procedure involved the additional steps in the early stages of heating strongly a further sealing-off point and of flushing out the vacuum branch with clean gas. The general procedure was equivalent to that described above up to the stage where the water sample was obtained. Then, sealing off at P_4 and at a point equivalent to P_8, the pump was used to exhaust the oxygen and a length of the tube S containing the sample could be drawn off. Successive condensations could then be performed and further lengths, containing samples, taken off the tube S.

Cooling experiments were also performed with water distilled directly into open tubes, no attempt being made to exclude the atmosphere. Six tubes of Hysil glass about 1 cm in diameter and 10 cm long, with one end sealed test-tube fashion, were prepared. These were cleaned with hot chromic acid and rinsed out several times with boiling doubly distilled water taken direct from a glass still. They were then filled with cooler water directly from the still.

All samples were cooled in the special Dewar vessel shown diagrammatically in fig. 4. Cooling was produced by the expansion of carbon dioxide gas from cylinder pressure to atmospheric pressure at a valve in the Dewar vessel, the expanded gas being allowed to pass through several turns of copper tube immersed in the alcohol medium. The alcohol was stirred briskly by bubbling nitrogen through it. The temperature was indicated by two toluene thermometers the errors of which were determined at 0°C, appropriate corrections being applied throughout the observations. Standard rates of cooling of 1 and 2 centigrade degrees per minute were used. Any lag error, which must in any case have been small, was consequently of no significance for relevant temperature comparisons. Because of the rapidity with which the observations had to be taken, when the consistency of the thermometers had been established by experience only one was used.

§5. RESULTS

The results showed clearly that the freezing process was governed by the glass surface. The following observations, which have some interesting features, are representative.

The freezing could occur successively for the same sample tube at a temperature which was almost constant or at temperatures which lay in a range and which were apparently governed by a probability distribution curve. These cases are illustrated by the behaviour of two samples which were drawn off after being evacuated of oxygen. The tubes were about 15 cm long and 1 cm in diameter and were almost half filled. One of these froze successively at the temperatures (taken to the nearest half degree) −12·4, −12·4, −11·9, −12·4, −11·4, −11·4 and −11·9°c. It was then inverted and froze at −7·9, −7·9 and −7·9°c to the nearest tenth of a degree. The other sample was frozen fifteen times and froze at temperatures ranging from −13·9 to −20·4°c in such a way as to suggest a probability distribution curve of the form shown in figure 5. The lowest temperature reached in the experiments was −30 ± 1°c, obtained twice with a sample produced in this way. The freezing temperature of this sample had lapsed on the following day to −21·4 ± 0·5°c and again the next day to −13·4 ± 0·5°c. Lapse of freezing temperature in this direction was the rule, but the effect was generally much smaller than this. Heating an enclosed sample by immersion in boiling water did not lower its freezing temperature.

Experiments with the S-shaped sample tubes showed that, when the sample of water was blown out by the oxygen and replaced by a further distillation, the freezing temperature and the site of nucleation were unaffected. The means of determining the nucleation site is described below (nucleation sites could not be ascertained for the furnace-heated tubes because colloidal graphite and asbestos which had been used to prevent welding of the softened glass to the furnace tube imparted an etched appearance to the tube). The following history of an S-tube shows a remarkably reproducible performance and also suggests that quite strong heating of the emptied tube does not necessarily damage the nucleation centre.

An S-shaped tube approximately 1 cm in diameter was cleaned as detailed above with acid, steam and firing, and when the tip had been sealed a few cubic centimetres of water were condensed into it from the oxygen stream. The first day freezing occurred from an observed point, which may be designated X, at temperatures −10·0 ± 0·5°c and −10·5 ± 0·5°c. The second day freezing occurred at −11·0 ± 0·5°c from the point X, and on the fourth day a value −11·5 ± 0·5°c was recorded, freezing again beginning at X. (A sample could be kept in an S-tube for any period without contamination simply by sealing the tip of E at P_4 and then opening the vessel W to the atmosphere at the nipple N. The oxygen stream could be restarted by closing N and then opening P_4 when a sufficient pressure of oxygen had developed.) This sample of water was then blown out with the oxygen and a further sample condensed as before. On the same day (fourth day) freezing temperatures −8·8, −9·0, −8·7, −8·4 and −8·4°c were obtained for this sample, freezing beginning at the point X within a few millimetres. On the fifth day this same sample gave freezing temperatures −8·5, −8·5 and −8·5°c, again freezing from the point X each time. It was refilled the same day, freezing temperatures −8·2, −8·4 and −8·4°c being obtained, the nucleation site being X.

The water sample was blown out and the glass was heated to incandescence in a small region on the opposite limb of the U-section to that bearing the point X. The tube was refilled and freezing temperatures −8·2, −8·2 and −8·0°c were recorded, freezing beginning from X. The tube was again emptied and the neighbourhood of thh point X itself was heated to incandescence (probably about 1000°c). It was refilled and freezing temperatures −9·7 and −10·2°c were obtained. However, freezing began at a point close to X, within the heated region, but not coincident with X. A further freeze gave a temperature of −11·0°c and indicated the original point X. The following day (sixth), the same sample gave −11·2°c, the site of nucleation not being very clear. The tube was emptied in the usual way. It was fired to a red heat generally, and particularly strongly in the neighbourhood of the active region. Freezing was then obtained at −8·5 and −9·5°c from an entirely new point on the opposite limb.

Experiments performed with the open tubes showed similar phenomena with somewhat less reproducibility. The nucleation sites, however, were more clearly evident than with the S-tubes. The freezing temperatures for tubes which had been numbered I, III and V are given in table 1. The asterisk indicates those observations immediately before which the tubes were rinsed out several times (from the still) and refilled. The dimensions, in inches, are the distances from the bottoms of the tubes of the points at which freezing began. These points could not be ascertained simply by watching the freezing occur, since the rate of propagation along the tube of the skeleton of crystals, distinguishable by the opalescent appearance, was roughly 10 cm/sec. However, they were usually discernible from the texture of the skeleton of crystals. This crystal mass consisted of a large number of needle-like dendritic crystals. In the neighbourhood of the nucleation site the crystals were directed radially away from a central point in a relatively conspicuous manner. Otherwise, where the crystals were dense, they appeared to be arranged in a more or less regular way. Where they were more sparse, such as in the neighbourhood which had corresponded to the level of the alcohol in the cooling vessel, they were observed to have formed by branching into beautiful fern-like sheets. These ice ferns, or at least portions of them, were accurately plane, for a specular reflection could be obtained from the whole of a 'fern leaf' at the same sharply defined angle. The symmetry of the branching showed that the hexagonal axis was perpendicular to the plane of the 'leaf'. The impression usually gained when the crystal mass spread suddenly and rapidly from the nucleation point was that of spiral or rather helical motion along the sample tube. A simple calculation shows that, for the freezing temperatures involved here, the immediate product is predominantly water interspersed with a smaller quantity of ice. The temperature in the tube rises immediately to 0°c and a solid layer of ice begins to grow slowly inward from the surface of the glass tube. When freezing was completed an obvious fissure usually remained along the tube axis, sometimes showing interference colours. The thawing of the ice was then accompanied by a considerable danger of cracking the tube. Consequently freezing was seldom allowed to go far beyond the initial stage of the ice skeleton.

Table 1. Freezing Temperatures (in °c) obtained with Open Tubes I, III, V

I	−16·0	−15·5	−15·2	−15·2	−15·2*	−15·2, 2½″	−14·8*, 2½″	−14·9*, 2½″
III	−14·2	−13·8	−14·7	−13·0	−14·0*, 1¼″	−12·8, 1¼″	−13·8*, 1¼″	−12·7*, 1¼″
V	−17·5	−17·6	−16·8	−16·9	−16·5*, ⅜″	−17·2*, top	−17·2*, ⅜″	−16·6, ⅜″

* Before taking observations marked with an asterisk the tubes were rinsed several times and refilled.

The distances, in inches, represent the positions of the observed nucleation sites relative to the bottoms of the tubes.

Advantage was taken of these six tubes to observe the nucleating ability of colloidal graphite, claimed by Smith-Johannsen (1948) to be effective at approximately −7°c, and of other forms of carbon. The temperature limit below which freezing could be initiated by scratching the insides of the glass tubes with a clean (Hysil) glass rod was also determined.

A little Aquadag was washed in distilled water by sedimentation and a suspension in distilled water, opaque at a thickness of one to two centimetres, was formed. When freezing temperatures had been recorded for clean water in the six tubes, one drop of suspension was added to each and the freezing temperatures again taken. The results are shown in table 2(*a*). In each case the freezing temperature was raised, the mean freezing temperatures being approximately −11·0°c. The figure given by Smith-Johannsen, which might have been of some meteorological importance, is thus not of general significance. Experiments with soot taken from a naphthalene flame and with charcoal prepared by firing a small quantity of sugar to a red heat, which were performed in other tubes, showed that these forms of carbon were no more effective in nucleating freezing. The glass tubes, still containing the water to which graphite had been added (this had settled at the bottom), were each cooled again, the interior surfaces being rubbed at intervals of a few tenths of a degree with a clean glass rod.

Table 2

(*a*) The Effect of Aquadag

Tube No.	I	II	III	IV	V	VI
Clean water (°c)	−14·6	−19·0	−12·0	−20·1	−15·5	−20·1
Clean water with Aquadag (°c)	−12·5	−11·9	−11·8	< −10	−11·4	−10·0

(*b*) The Effect of Abrasion (Glass on Glass)

Tube No.	I	II	III	IV	V	VI
With abrasion (°c)	−2·5	−2·5	−2·2	—	−3·0	−2·5
Without abrasion (°c)	−12·5	−11·8	−12·2	—	−9·3	−9·1

(*c*) The Effect of Dust

Tube No.	1	2
Clean water (° c)	−10·2	−18·6
Clean water with dust (°c)	−5·5	−6·3
,, ,, ,,	−5·3	−5·5
,, ,, ,,	−5·3	−5·1

The samples were then frozen again without rubbing. The results, given in table 2(*b*), show that rubbing gave rise to nucleation at a temperature of approximately −2·5°c, and that after this treatment the initial freezing temperatures (i.e. the final temperatures of 2(*a*)) were approximately reproduced.

Two tubes were cleaned and filled with distilled water as above and, when their freezing temperatures had been recorded, a little room dust, taken from the top of a cupboard and presumably representative of the colloidal matter of the atmosphere, was added to each. The subsequent freezing observations are shown in table 2(*c*), the mean freezing temperature being −5·4°c.

§6. CONCLUSIONS

When contamination by the colloidal matter of the atmosphere was prevented, the freezing occurred for a particular glass tube in a remarkably reproducible manner, often as regards both the location of the site at which nucleation occurs and the temperature. It is natural to enquire at what kind of singularity initiation occurs.

It is known that the surface of a glass can be profoundly modified by the action of water, acids and alkalis. Further, in borosilicate glasses such as Hysil prolonged heating at temperatures of 800° to 1000°c can cause visible crystals of tridymite to form on the surface. This being so, the cleaning and heating of the container S might well have resulted in the formation of occasional minute crystals on the glass surface. It would not be surprising if such crystals were capable of nucleating the freezing of water at -10°c or so.

This view is supported by the facts that the nucleation sites were unaffected by boiling water and not always affected by treatment with acid or by quite strong heating. The technique of freezing water in cleaned glass vessels is not likely to lead very far, for there is little prospect that glasses other than that used would behave very differently. It might be worth while to perform similar experiments in pure silica glass when the only possible devitrification products could be tridymite and cristobalite (and hypothetically quartz, which has, however, never been observed as a devitrification product of any glass), and when much higher temperatures could be used in cleaning. The possibility exists that, in this case, not only the behaviour of an individual tube would be reproducible, but that a large measure of similarity of behaviour would be observed for different tubes.

An outstanding feature is apparent when the close reproducibility of freezing temperature in the above results is considered in relation to Frenkel's theory as represented by fig. 1. It is seen from the figure that, unless the shape of the curves departs greatly from the true shape, even for a nucleation process in which the free energy of activation is only a small fraction of that corresponding to nucleation in the bulk of the liquid, the freezing temperature, often reproducible within limits as close as $\pm 0{\cdot}2$°c, must lie on the steep part of the appropriate curve and between about -5 and -15°c. This means that if the curves of fig. 1 were re-drawn with the ordinate on a linear scale, the freezing temperature would fall between the melting point and the apparent toe of the appropriate curve. In other words, for the size of sample used, the probability barrier which still lies ahead at the instant of freezing is so great that an entirely negligible chance exists that the temperature could be lowered to the region of low freezing probability without freezing occurring at an intermediate stage.

On the question of whether atmospheric contamination was a significant factor in the experiments carried out in open tubes, the considerable difference in consistency between the results for open and closed tubes strongly suggests that it was. Further, the ability of room dust to cause nucleation at about -5°c shows that freezing at this temperature due to airborne matter could be expected occasionally, but it is not known with what probability. All that can be stated categorically is that under the conditions of the experiments the atmosphere did not affect the behaviour of the exposed water with high probability for temperatures above about -16°c. Consideration of all results obtained showed that the *a priori* probability of freezing temperatures for the glass tubes was negligible for temperatures above about -6°c, a maximum at -10 to -12°c, dropping off rapidly between -18 and -20°c and becoming negligible at about -30°c.

The tests using forms of carbon require little comment. Particular interest attaches to carbon as a nucleating material since it is known to be present in the atmosphere. However, the nucleating ability observed in the above experiments is no greater than has been claimed for several other substances, although the Aquadag clearly had a positive effect.

The effect of friction between two glass surfaces is reminiscent of the effect of mechanical violence in inducing other phase changes such as crystallization from solution. The experiments showed that no permanent nucleating ability was bestowed on the abraded glass surface. Friction, or at least mechanical action, provides the only means at present known (other than seeding by ice) of causing freezing in supercooled water at temperatures above about -5°c. Young and Van Sicklen (1913) found that water could be induced to freeze by the very violent impact of a hard steel point on a steel surface. These workers, by forcing the freezing temperature well above -5°c, were able to find a well-defined relationship between the violence of the impact and the highest temperature at which it was effective in producing freezing. Their most violent impacts initiated freezing at temperatures above $-0{\cdot}1^\circ$c.

Freezing produced by mechanical means has usually been explained in rather loose terms by reference to 'shock' and to the disturbance of oriented surface layers. However, following up the suggestion of the Earl of Berkeley (1912) that nucleation is caused by the pressure waves produced, it is not a difficult matter to show that the effects of friction or impact can be explained as being due to large local changes in pressure. Here we may first estimate the change in pressure necessary to produce the observed effects and then show how the requisite large extensive pressures can occur. Since the result can, in any case, be correct only in order of magnitude, the Clausius–Clapeyron equation will be sufficiently accurate even though the increments of temperature and pressure are large. In other words, the dependence of the mutual latent heat and the molecular volumes of water and ice on temperature and on pressure will be disregarded. The change in melting point δT due to a change in pressure δP is then given by $\delta T/\delta P = T(v_\alpha - v_\beta)/L$ with the same notation as before, L being positive. Now, assuming that a supercooling of the order of 50°c will ordinarily cause true self-nucleation with a very high probability (cf. fig. 1), it is logical to expect that a pressure change which alters the melting point from 0 to $+50^\circ$c will cause the onset of freezing with high probability at, say, -1°c. The pressure change which will do this is a large decrease in pressure, so that negative or extensive pressures are involved. The Clausius–Clapeyron equation gives a value for δP of approximately -5000 atmospheres when $\delta T = 50^\circ$c. The question is, then, can extensive pressures of the order of 5000 atmospheres develop during friction or impact in water?

In this connection it is desirable first to determine what is the maximum extensive pressure that water can withstand without self-cavitation. Following Frenkel (1946) and using the simplest theory, which assumes that the required extensive pressure is that which will equilibrate a bubble whose diameter is equal to the intermolecular distance in water, a figure of 10 000 atmospheres is obtained. Attainment of the pressures required by the above theory is thus not precluded by self-cavitation, at least at constant temperature. The estimation of the local pressure changes which result when the rubbing together or impact of solids occurs in a liquid is not easy, especially when only negative pressures are of interest. Large extensive pressures can presumably arise in two ways, namely from the rapid separation of solid surfaces and from the elastic rebound which must follow a localized compression, particularly a compression due to the collapse of a cavity.

When two glass surfaces are abraded together, local welding followed by tearing is to be expected, the damage to the surfaces being evidenced in the form of scratches. Very large local accelerations must be involved in the disruptive process, and it seems probable that cavitation will occur, even if it begins at the solid surfaces, for a negative pressure considerably smaller than that required for self-cavitation in the liquid. If a cavity of any magnitude is formed its collapse will probably be associated with an elastic oscillation involving very large pressure changes. The determination of the maximum pressure reached, which is of importance also in the study of underwater explosions, is not a simple matter. The collapse of a spherical cavity filled with a permanent gas at low pressure gives rise to enormous pressures; in fact, if the initial pressure in the cavity is p_0 and the external pressure is P then, neglecting viscosity and the compressibility of the liquid, the energy equation for isothermal collapse gives for the maximum pressure $p=p_0 \exp(P/p_0)$. When P is one atmosphere any reasonable value of p_0 gives an enormous value for p. For example, $p_0=20$ mm Hg corresponds to $p \sim 10^{-14}$ atmosphere. In practice the cavity is filled with vapour, the condensation of which is impeded by the liberation of latent heat. The effect is to damp the motion. On collapse, the cavity will probably disappear altogether, leaving a radial pressure distribution which will cause the liquid to move outward elastically; an extensive pressure will develop at the centre. The actual pressures are not likely to be as large as the value estimated for a gas-filled cavity, but it is not at present possible to say what is their actual magnitude In a recent theoretical treatment of cavitation, Neppiras and Noltingk (1951) arrived at values of the order of 10^6 atmospheres for the maximum collapse pressures of vapour cavities.

The result obtained by Maurin and Médard (1947) for the freezing of supercooled water clouds subjected to shock waves suggests that much smaller pressures were effective in their circumstances, but the possibility exists that some dynamical process associated with the droplets gave rise to localized amplified pressure effects and, of course, solid particles were very probably present in the cloud drops.

Cavities which form in practice will not in general be spherical, on which account they will be considerably less effective in producing large pressure changes. On the other hand they may collapse on to the glass surface, on which account their effectiveness in causing nucleation is probably enhanced.

In the light of these considerations, whilst it has not been possible to determine conclusively the order of magnitude of the pressure fluctuations which accompany the mutual impact and abrasion of solids, it seems very probable that large instantaneous negative pressures are responsible for the ability of violent mechanical processes to nucleate freezing in supercooled water.

ACKNOWLEDGMENTS

The work described above was carried out at the H. H. Wills Physical Laboratory, University of Bristol. The author wishes to express his appreciation to Dr. F. C. Frank of that laboratory for his continuous interest and for helpful discussions, and to thank Professor H. Moore of the University of Sheffield for having drawn his attention to some known properties of borosilicate glasses.

REFERENCES

BECKER, R., and DÖRING, W., 1935, *Ann. Phys., Lpz.*, **24,** 719.

BERKELEY, Earl of, 1912, *Phil. Mag.*, **24,** 254.

DORSEY, N. E., 1948, *Trans. Amer. Phil. Soc.*, **38,** Pt. 3, 246.

EINSTEIN, A., 1910, *Ann. Phys., Lpz.*, **33,** 1275.

FOURNIER D'ALBE, E. M., 1948, *Nature, Lond.*, **162,** 921.

FRANK, F. C., 1949, *Discussions of the Faraday Society*, **5,** 48.

FRENKEL, J., 1932, *Sowjet Phys.*, **1,** 498; 1946, *Kinetic Theory of Liquids* (Oxford : University Press).

LAFARGUE, C., 1950, *C. R. Acad. Sci., Paris*, **230,** 2022.

MAURIN, J., and MÉDARD, L., 1947, *C.R. Acad. Sci., Paris*, **225,** 432.

MEYER, J., and PFAFF, W., 1935, *Z. anorg. Chem.*, **224,** 305.

NEPPIRAS, E. A., and NOLTINGK, B. E., 1951, *Proc. Phys. Soc.* B, **64,** 1032.

SANDER, A., and DAMKOHLER, G., 1943, *Naturwissenschaften*, **31,** 460.

SMITH-JOHANNSEN, R., 1948, *Science*, **108,** 652.

TAMMANN, G., and BUCHNER, A., 1935, *Z. anorg. Chem.*, **222,** 371.

THOMSON, J. J., 1888, *Applications of Dynamics to Physics and Chemistry* (London : Macmillan).

TURNBULL, D., and FISHER, J. C., 1949, *J. Chem. Phys.*, **17,** 71.

VOLMER, M., 1939, *Kinetik der Phasenbildung* (Leipzig : Steinkopff).

VOLMER, M., and FLOOD, H., 1934, *Z. Phys. Chem.* A, **170,** 273.

VOLMER, M., and WEBER, A., 1926, *Z. Phys. Chem.*, **119,** 277.

WILSON, C. T. R., 1897, *Phil. Trans. Roy. Soc.* A, **189,** 265.

YOUNG, S. W., and VAN SICKLEN, W. J., 1913, *J. Amer. Chem. Soc.*, **35,** 1067.

LETTERS TO THE EDITOR

The Saturation Magnetization of Nickel at High Pressure

Observations of changes in ferromagnetic behaviour under pressure have been made by various authors, including Ewing (1900), employing linear compression, and Chi-Sun Yeh (1925) and Ebert and Kussmann (1937), employing hydrostatic compression, though only the latter authors have used sufficiently high fields to reach saturation. The changes in the saturation magnetization I_{sat} observed by them were so small as to amount to negative results for most of the ferromagnetics used (including nickel) but it seems to us that at high fields their apparatus must have acted to keep the flux constant in such a way that small changes in the susceptibilities of specimens would not have been detected.

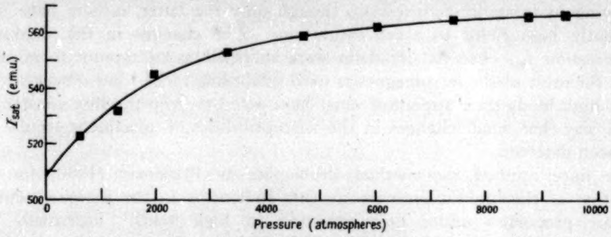
We have applied the method developed by Bridgman (1950) for the generation of the highest pressures in thin specimens to the measurement of magnetic properties under high pressures in high fields. Essentially our apparatus consists of an electromagnet in which truncated conical pole-pieces are used both to transmit the force applied by a hydraulic press to thin specimens held between them and to concentrate the magnetic flux into the specimen. Where I_{sat} for the specimen is not too large the use of very thin specimens enables high fields to be applied. The stress in the specimen is not purely hydrostatic, being intermediate between hydrostatic and linear pressure. The values of pressure quoted below are, strictly, values of the force applied per unit area of specimen.

By measuring separately the flux through the specimen and through the air between the pole-pieces for several thicknesses of non-magnetic specimen, effects due to changes in other parts of the magnetic circuit are eliminated, and (I, H) curves can subsequently be obtained for the material of any other specimen. Pressures up to 80 000 atm can be applied, using existing materials, although the magnetic hardness of the strongest materials (such as cobalt steel) greatly reduces the sensitivity of measurement at such pressures. Using mild steel pole-piece tips, pressures up to 10 000 atm and fields up to 20 000 oersteds can be applied.

We have determined (I, H) curves for nickel up to 10 000 atm, at which pressure the internuclear distance is decreased by about 0·5%. At all fields magnetization increases with pressure, as found by Ewing and by Chi-Sun Yeh. We have, however, been able to extend measurements to complete saturation, using fields up to 12 000 oersteds, and these results are summarized in the figure. The curve represents the reversible behaviour of work-hardened nickel. The values of I_{sat} obtained, though consistently high by about 2%, are believed to be accurate relatively to about 0·3%.

Changes in I_{sat} directly reflect changes in the electronic band structure, and these results imply, for nickel, an increase by 11% at the highest pressures in the number of unoccupied 3d states (normally 0·6 per atom) and the number of occupied 4s states. It is to be expected that both the 3d and 4s bands will be broadened by compression. However, in order to predict the influence of pressure upon I_{sat} it would have been necessary to know with some precision the form of the curve of density of states against energy for both bands as a function of atomic spacing. It is not possible, conversely, to make more than a plausible guess at this by considering the implications of our results.

We have also carried out tests up to 10 000 atm on the following non-ferromagnetic metals which might conceivably be expected to become ferromagnetic at high positive or negative pressures: Pd, Pt, Mo, W, Cu, Ag, Au.



In each case the volume susceptibility was shown not to exceed 0·0003 at high fields and 0·0006 at low fields. Of interest in itself, this negative result is a useful check on the accuracy of the method.

A full account of this work will be published later. We gratefully acknowledge valuable suggestions and help given by Dr. J. R. Barker of this Department and Mr. F. F. Roberts of the Post Office Engineering Department, and support given by the Department of Scientific and Industrial Research and the Central Research Fund of the University of London.

Department of Physics,
Queen Mary College,
University of London.
3rd December 1952.

G. O. Jones.
F. D. Stacey.

Bridgman, P. W., 1950, *Proc. Roy. Soc.* A **203,** 1.
Chi-Sun Yeh, 1925, *Am. Acad. Arts and Sci. Proc.,* **60,** 503.
Ebert, H., and Kussmann, A., 1937, *Z. Phys.,* **38,** 437.
Ewing, J. A., 1900, *Magnetic Induction in Iron and Other Metals* (London: 'Electrician' Publ. Co.).

The 'Corona Pressure' and Negative Joshi Effect

Joshi postulated (1943, 1946, 1947) that the sharp reversible decrease in current $-\Delta i$ in a gas under discharge when exposed to external light of any frequency, from extreme red (Joshi 1945) to x-rays (Joshi 1944) results from a space charge effect leading to a lowering of the effective mobility k. A new way of testing this postulate appears to be by measuring the *corona pressure*, i.e. the almost instantaneous reversible pressure rise Δp due to a discharge, under conditions when the Joshi effect is simultaneously observed. For wire-cylinder coronae, Tyndall and Searle (1918) showed that

$$\Delta p = i/4\pi k \qquad \ldots\ldots(1)$$

where i is the discharge current per unit electrode length and k the ionic mobility. p, i and k would be expected to be similarly related in other discharges, and hence measurements of corona pressure in the dark Δp_D and under illumination Δp_L may be expected to reveal if $-\Delta i$ is associated with a decrease in k. Such measurements have been made for air, oxygen, nitrogen and hydrogen in different types of discharges over a wide range of operating conditions (Arnikar 1949). The table shows a typical set of results for pure dry hydrogen in a low-frequency ozonizer discharge irradiated by light from a 200 watt glass incandescent lamp. A filter of alum solution was used to minimize the heating effect of the lamp. Δp was measured with a mercury manometer read by a telescope with an eye-piece scale graduated in units of $\frac{1}{40}$ mm.

Potential kv (r.m.s.)	Corona pressure* Δp_D	 Δp_L	Current† i_D	 i_L
	(i) $p = 242$ mm Hg			
1·05	2	2	65	39
1·40	2·5	2·5	85	47
1·74	4	4	110	72
2·08	5	5	135	88
	(ii) $p = 302$ mm Hg			
1·05	3	—	62	37
1·40	5	5	80	51
1·74	6	8	115	80
2·08	9	10	140	102
	(iii) $p = 370$ mm Hg			
1·05	3	4	60	47
1·40	6	6	80	57
1·74	7	9	120	104

*** In units of $\frac{1}{40}$ mm Hg.** **† In arbitrary units of galvanometer deflection.**

Errors in Δp might arise from (*a*) temperature changes of the gas due to heat from the light source, (*b*) variation in applied potential, and (*c*) errors in reading the telescope. The results can be regarded as sensibly free from these errors, because (*a*) no pressure change could be observed before about 2 minutes with the *unexcited* tube exposed to light, whereas measurements of Δp were completed within 30 seconds, (*b*) series of observations showed that fluctuations of potential were unappreciable in the observation interval of 30 seconds, (*c*) pressure changes of $\frac{1}{80}$ mm could be estimated.

The results show that Δp is the same with illumination as in the dark, or increases slightly. No other pressure effect in gases like hydrogen due to irradiation by visible light is known. Hence, since the current decreases on illumination, it is suggested on the basis of eqn. (1) that the fact that Δp *does not decrease* lends support to Joshi's view that $-\Delta i$ is associated with a decrease of k.

Agashe (1952) has recently reported similar results. He assumes that metastable molecules are present which dissociate on irradiation, and that the increase in Δp which he found is due to this increase in the number of particles. At least in the case of hydrogen, which does not have metastable states (Loeb 1939), such an assumption is unwarranted, nor is it necessary on Joshi's theory.

Grateful thanks are due to Professor Joshi for suggesting the problem, and for valuable advice and guidance.

Chemical Laboratories,
Benares University,
Benares, India.
27th November 1952

H. J. ARNIKAR.

AGASHE, V. V., 1952, *Proc. Phys. Soc.* B, **65,** 740.

ARNIKAR, H. J., 1949, *Proc. Indian Sci. Congress Phys. Section, Abstract No.* 15.

JOSHI, S. S., 1943, *Presidential Address to the Chemistry Section, Indian Science Congress*; 1944, *Current Science*, **13,** 278 ; 1945, *Ibid.*, **14,** 317 ; 1946, *Proc. Indian Sci. Congress Phys. Section, Abstract No.* 26; 1947, *Ibid., Abstract No.* 25.

JOSHI, S. S., and DESHMUKH, G. S., 1941, *Nature, Lond.*, **147,** 806.

LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York: Wiley), p. 496.

TYNDALL, A. M., and SEARLE, N. S., 1918, *Phil. Mag.*, **35,** 261.

Sensitive Bubble Jets

During a study of bubble formation by rapid air flow (40–730 cm^3/sec) through rectangular slots submerged in water (Spells and Bakowski 1950), a few experiments were carried out to determine whether the bubbling process was affected by sound.

A 22-mm bore copper tube, 1 mm wall thickness, was set up vertically with its open lower end dipping into a trough of water. The bubbling was from a vertical slot, 3 mm wide × 25 mm, cut in the end of the tube; depths of immersion, to the top of the slot, were 3 cm or 3·5 cm. The tube, whose effective length could be varied from 24 cm to 41 cm by means of a telescopic joint, projected down from the horizontal floor of an airtight vessel of about 35 litres capacity connected to the air supply. Rigidly mounted in the vessel was a 'Vitavox' 20 w loudspeaker having a 12 in. diaphragm. The loudspeaker faced vertically downwards, being set coaxially with the slotted bubbling tube, with the diaphragm rim about 6 in. above the vessel floor, and was driven by a Dawe Wide Range Oscillator (20 c/s to 20 000 c/s), Type 400A, and amplifier. The air, then, flowed from the vessel containing the loudspeaker down through the tube, whence it bubbled into the water from the slot at the bottom.

Visual observation sufficed to show that the bubbling was affected at frequencies corresponding to first order resonance in the air column of the slotted tube. Obvious effects were also obtained at a higher frequency (1070 c/s), independent of the tube length; this frequency was probably associated with some other part of the apparatus.

Instantaneous flash photographs were taken using a Siemens 'Sieflash' tube. Figures 1 (slot edge-on) and 2 (slot facing camera) (Plate) show the effects of frequencies of 295 c/s (resonance in tube) and 1070 c/s respectively. Figures 1 and 2 should be compared with fig. 3 which, from previous work using high-speed cinematography (Spells and Bakowski 1950), is known to be typical of the appearance of the unserenaded bubbles. A high-speed ciné camera was not available when satisfactory conditions had been established for demonstrating the effect of sound, but evidently the bubbles are appreciably altered both in general shape and superficial roughness.

These experiments were inspired by the account by Smith (1935) of the resonant bubble, in which it was found that gas bubbles in water have characteristic frequencies depending on bubble volume and gas and liquid properties. Since, however, the present conditions, in which the bubble volumes change rapidly, are very different from Smith's, it is doubtful whether the effects observed are due to bubble resonance. Notwithstanding complications due to there being both a gas and a liquid phase, it is therefore suggested tentatively that these effects may belong to the well-known group of phenomena, embracing the sensitivity to sound of various gas-into-gas or liquid-into-liquid jets, described collectively by the term 'sensitive flame'. The shape of the air–water interface in fig. 2, suggestive of the alternate vortex formation observed by Brown (1935) and by Andrade (1941), in sensitive flame experiments, gives some support for this view; but confirmation must await more thorough investigation by cinematography.

The arrangement used, being dictated largely by the requirements of another investigation, is probably not the most effective for demonstrating the phenomenon. It is hoped that these rather inconclusive experiments may lead to further investigations being carried out elsewhere.

I am indebted to Dr. S. Bakowski and Miss M. Warburton of this Laboratory for taking the photographs from which figs. 2 and 3 were prepared.

Imperial Chemical Industries Ltd.,
General Chemicals Division, Research Department,
Widnes, Lancs.
6th November 1952.

K. E. SPELLS.*

* Now at R.A.F. Institute of Aviation Medicine, Farnborough, Hants.

ANDRADE, E. N. DA C., 1941, *Proc. Phys. Soc.*, **53**, 329.
BROWN, G. B., 1935, *Proc. Phys. Soc.*, **47**, 703.
SMITH, F. D., 1935, *Phil. Mag.*, **19**, 1147.
SPELLS, K. E., and BAKOWSKI, S., 1950, *Trans. Instn. Chem. Engrs.*, **28**, 38.

Noise in Semiconductors at Very Low Frequencies

It is well known that excess noise is observed when current is flowing in a semiconductor and the mean square noise voltage per unit bandwidth is approximately inversely proportional to frequency over a considerable range of frequencies (Montgomery 1952). From consideration of the mechanisms which might be responsible for the noise, it seems improbable that the $1/f$ law is valid at very low or very high frequencies. The theories which have been proposed (Macfarlane 1950, Richardson 1950, van der Ziel 1950, du Pré 1950, Surdin 1951, Petritz 1952) lead to different results for the frequency dependence at very low frequencies. In previous measurements (Kronenberger 1951) it has been found that the $1/f$ law is obeyed down to frequencies of the order of 0·2 c/s.

We have recently developed a method by which measurements can be made at much lower frequencies and have obtained preliminary results for the noise in pyrolytic carbon resistors. The principal feature of the method is that the noise is recorded on magnetic tape running at a very slow speed and the frequency analysis is then made by running the tape through the pick-up head at the normal speed. In this way all frequencies in the original record are multiplied by a known factor and the spectrum can be analysed with a standard type of audio-frequency wave analyser.

The record occupied a length of tape of about 18 in. which was joined into a closed loop. With a single recording a frequency range of more than two decades could be investigated. By varying the recording speed, measurements were made over a range of about five decades.

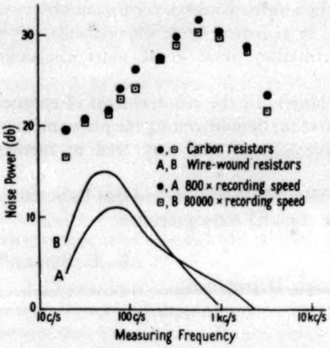


Fig. 1. Noise as measured on wave analyser.

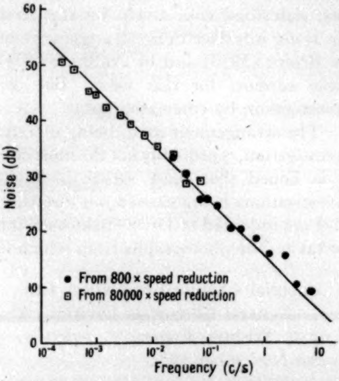


Fig. 2. Corrected results: Graph of noise-power per unit bandwidth plotted against frequency. The continuous line corresponds to $\overline{\Delta R^2}/R^2 = 10^{-13}\, df/f$.

Considerable amplification of the noise was required in order to obtain a level suitable for recording. The difficulties associated with amplification at very low frequencies were avoided by obtaining the noise as a modulation on an audio-frequency carrier. The resistors were incorporated in a bridge circuit operated at a frequency of about 1 kc/s. Fluctuations in the output from the bridge, due to fluctuations of the resistors, were amplified by a tuned amplifier and rectified by a phase sensitive linear detector. To avoid spurious effects due to temperature variations, the temperature of the bridge circuit was thermostatically controlled. A check was made to ensure that there was no appreciable distortion of the noise spectrum due to non-linearity of the recording system.

The results obtained from the analysis of two different records are shown in fig. 1. The ordinate is proportional to the power output from the pick-up head for a constant bandwidth of 4 c/s. In this figure no correction has been made for the playback characteristic of the system which, for a constant level recording, would give a voltage output proportional to frequency at low frequencies with a drop at high frequencies due to the finite width of the gap in the head. For noise proportional to df/f, recordings made at different speeds should be indistinguishable and should give the same output. It will be seen that this is indeed the case for the two recordings shown. For comparison, recordings made with wire wound resistors in place of carbon resistors are included in the figure to show the magnitude of effects due to spurious causes.

After the recording speed and playback characteristics have been taken into account, the results are as shown in fig. 2. It will be seen that the two recordings overlap satisfactorily and that, even at the lowest frequency, the $1/f$ law is still fairly closely obeyed.

We hope to continue this work with measurements of the noise in single crystals of silicon and germanium and in rectifying contacts at different temperatures and a full account will be given at a later date.

The Clarendon Laboratory,
Oxford.
B. V. ROLLIN.
I. M. TEMPLETON.

13th December 1952.

KRONENBERGER, K., 1951, *Z. angew. Phys.*, **3**, 1.
MACFARLANE, G. G., 1950, *Proc. Phys. Soc.* B, **63**, 807.
MONTGOMERY, H. C., 1952, *Bell. Syst. Tech. J.*, **31**, 950.
PETRITZ, R. L., 1952, *Phys. Rev.*, **87**, 535.
DU PRÉ, F. K., 1950, *Phys. Rev.*, **78**, 615.
RICHARDSON, J. M., 1950, *Bell. Syst. Tech. J.*, **29**, 117.
SURDIN, M., 1951, *J. Phys. Radium*, **12**, 777.
VAN DER ZIEL, A., 1950, *Physica*, **16**, 359.

REVIEWS OF BOOKS

Colour in Theory and Practice, edited by H. D. MURRAY. Pp. xiii + 360. (London: Chapman and Hall, 1952.) 70*s*.

Boyle, Newton, Young, Dalton, Goethe, Maxwell, Rayleigh, Schroedinger; it is curious how many men, whose chief claims to fame lie in other realms, have included colour amongst their researches. Perhaps its fascination springs largely from the fact that colour is an everyday phenomenon which can only properly be understood when several normally separate disciplines are brought together. Every time we see a colour, a *physical* stimulus, perhaps modified by *chemical* absorption, excites a *physiological* organ, to produce a *psychological* sensation. It is not, therefore, surprising to find that no fewer than eight authors have contributed to *Colour in Theory and Practice* which, although in one sense a new edition of the volume by Murray and Spencer published under the same title in 1939, is really substantially a new work. Four physicists, two chemists, one physiologist, and one electrical engineer form the present team, and the result is a volume containing a great deal of very useful information, many valuable diagrams and tables (including a good selection of those which have appeared to such good advantage in recent volumes of the Journal of the Optical Society of America), and the whole very well printed and produced.

The book is in four main sections the first of which deals with physical and chemical aspects of colour and serves to introduce the physicist to the appropriate chemistry and the chemist to the appropriate physics. The treatment, from this point of view, strikes an excellent balance, being intelligible, concise, and including fundamentals of emission and absorption processes.

In the second section, physiological and psychophysical aspects are described, and here, quite rightly, the field narrows and the treatment deepens. On the physiological side, the processes of normal and defective colour vision are explained in so far as they are known, and the psychophysics of colour matching, measuring, and specification complete the section. The explanation given for the tendency of red colours to appear to stand out in front of their actual location seems doubtful since some observers insist that, to them, red colours recede, and the text and figure 10.12 do not seem to tell quite the same story, but these are small blemishes on a well-balanced treatment. The chapter on colour measurement is admirably clear and concise and makes good use of geometrical arguments, although the experimental facts upon which they ultimately rest are not always too well emphasized. Thus the validity of the process of integration described on page 134, which is so widely used in colorimetry, is dependent on the validity of the Law of Additive Mixture, but this is not stated. The C.I.E. primaries, X, Y, and Z are described as imaginary, but they can, nevertheless, be used in a visual tricolorimeter.

The treatment of colour systems and tolerances handles two difficult subjects extremely well, but it is a great pity that one of the colour plates could not have been an approximate rendering of a page from, for example, the Munsell Colour Atlas. In this connection also, one would have liked to have seen an example of a Colour Confusion Chart as another colour plate, while the plates bearing figures 10.9–10.11, and figure 16.7 could certainly be dispensed with, if necessary.

The title of the third section is Light Sources and Colorimetry. Considerable development has taken place in the production of new types of electric lamp during the last two decades, and it is well summarized in this section, which is copiously illustrated. The principles of spectrophotometry and colorimetry, together with descriptions of the more important types of instrument used in these fields are, on the whole, well described. It is particularly valuable to have the table of selected ordinates reprinted from Hardy's *Handbook of Colorimetry*, and the worked examples of both the weighted, and the selected, ordinate methods are an excellent feature.

A brief final section, Miscellaneous Aspects, describing some of the work done on the recognition of 'warm' and 'cool' colours, colour preference, and colour in nature, rounds the book off in a very acceptable manner.

The three appendices provide a short catalogue of important colouring substances, eighteen pages of very valuable colorimetric tables, and a substantial extract from the very useful *Report on Colour Terminology* of the Physical Society Colour Group. Good bibliographies are given at the end of each chapter, and an adequate index is included.

It is unfortunate that, having praised many of the excellent features of this book, it has to be added that there is a curious lack of polish in the work. For instance, on page 25, Planck's formula is incorrect ; in figure 3.7 the wavelength scale is out ; in figure 10.6, a and b should be interchanged in order to agree with the text; in figure 11.7 (*a*) R. G. B. should be X, Y, Z; in chapter 14 two figures have been omitted altogether and meaningless reference is made to figures 14.2 and 3 instead ; on page 202 the figure referred to is 3.11 and not 3.12 ; on page 249 figure 4 should read figure 18.3 ; the description on page 284 of figure 20.7 is so brief as to be quite unintelligible, no reference being made to any of the 22 letters upon it ; and figure 21.3 contains 3 types of dot, but only two are described. On page 312 the formula of cobalt yellow should be $K_3Co(NO_2)_6$.

Although photometry may be regarded as outside the scope of a book on colour it seems a pity that in a volume, otherwise so complete, no definition of luminance is given except in an appendix, and that ' stilbs ' and ' foot-candles ' are used without any explanation.

Any book on colour will be judged by some readers largely on the way in which the C.I.E. system of colour measurement is explained. Admiration will be felt for the effort made in chapter 11 to make things clear, but it is unfortunate that a notation officially deprecated by the C.I.E. has been used; and in chapter 18 equations of the type $X+Y+Z=E$ have been used in a sense entirely different from that commonly found in American literature; and on page 272 another type of colour equation altogether is used. Incidentally, on page 135 it is rather disconcerting to read that a colorimeter employing more than three matching stimuli would not be worth while, when the many advantages of the Donaldson 6-primary instrument are very properly enumerated on page 259.

Colour in Theory and Practice is a book, however, to be welcomed, because it treats the subject very comprehensively, and includes much information hitherto only available from widely scattered sources. If its blemishes occasionally irritate the reader, let him regard them as the price which, it seems, has to be paid for a work contributed to by no fewer than eight writers, without whose participation the broad authoritative nature of the book could not have been achieved.

R. W. G. HUNT.

Instruments and Accessories for Radio-Isotope Applications. Pp. 17. (London: The Scientific Instrument Manufacturers' Association, 1952.) Copies free on application to the Association.

Séries de Fourier régularité, séries divergentes et formulation expérimentale, by PIERRE VERNOTTE. Pp. xvi + 105. (Paris : Publications Scientifiques et Techniques du Ministère de l'Air, 1952.) 800 fr.

CONTENTS OF SECTION A

PAGE

Mr. J. F. Nicholas. Effect of the Fermi Energy on the Stability of Superlattices . 201

Dr. L. H. Sutcliffe and Dr. A. D. Walsh. The Ultra-Violet Absorption Spectrum of Nitric Oxide 209

Dr. B. B. Goodman. The Thermal Conductivity of Superconducting Tin below 1°K 217

Dr. R. W. Hill and Mr. P. L. Smith. The Anomalous Specific Heat of Ferrous Ammonium Sulphate 228

Mr. G. R. Allcock. Damping Corrections in the Photo-Meson Process . . 233

Mr. P. Swan. The Elastic Scattering of Neutrons by Tritons at 14 Mev . . 238

Dr. J. R. Holt and Mr. T. N. Marsham. An Investigation of (d, p) Stripping Reactions—I : Apparatus and Results for Aluminium 249

Dr. J. R. Holt and Mr. T. N. Marsham. An Investigation of (d, p) Stripping Reactions—II : Results for the Isotopes of Magnesium 258

Dr. B. H. Bransden and Dr. A. Dalgarno. The Application of Variational Methods to Scattering by Ions—I : The Elastic Scattering of Electrons by Helium Ions 268

Dr. B. N. Srivastava and Mr. M. P. Madan. Thermal Diffusion of Gas Mixtures and Forces between Unlike Molecules 278

Dr. I. Abdelnabi and Prof. H. S. W. Massey. Inelastic Collisions of Electrons in Helium and Townsend's Ionization Coefficient 288

Mr. A. J. Salmon and Mr. E. K. Inall. The Angular Correlation of the Protons and γ-Radiation from the Reaction ^{6}Li(d, p)^{7}Li$^*\gamma$^{7}Li 297

Research Note :

Dr. J. Tunstead. Photoelectric Absorption in Lithium Vapour 304

Letters to the Editor :

Mr. J. M. Baker, Dr. B. Bleaney, Mr. K. D. Bowers, Dr. P. F. D. Shaw and Mr. R. S. Trenam. Nuclear Spin and Magnetic Moment of Radioactive Cobalt 57 305

Mr. D. West. Measurements of the Energy Loss Distribution for Minimum Ionizing Electrons in a Proportional Counter 306

Dr. L. Singh. On the Effect of a Strong Electrostatic Field on Scattering. 309

Mr. R. G. Jarvis and Dr. D. Roaf. The D–^{3}H and D–^{3}He Reactions below 45 kev 310

Reviews of Books 310

Contents of Section B 312

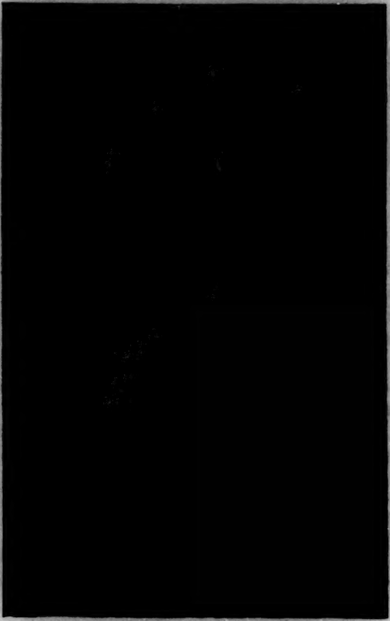


Fig. 1. 200 cm³/sec (295 c/s) (side view).

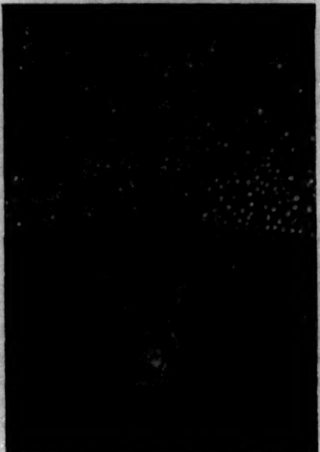


Fig. 2. 84 cm³/sec (1070 c/s).

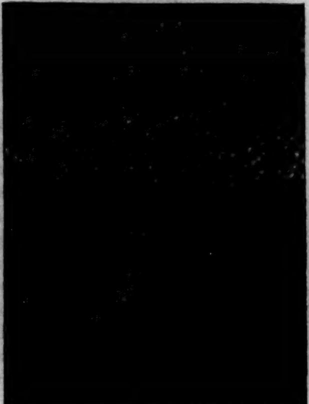


Fig. 3. 84 cm³/sec (silent).

An Interpretation of the Magnetic Properties of some Iron-Oxide Powders : II

By W. P. OSMOND

Magnetic Tapes Division, Minnesota Mining and Manufacturing Co., Ltd., Slough, Bucks.

MS. received 9th January 1953

Abstract. An earlier study of the magnetic properties of fine dispersed powders of acicular grains of magnetite and of γ-ferric oxide, derived by reduction and re-oxidation of dehydrated pigment oxides, has been revised, taking account of the probable cavities in the powder particles caused by the loss of oxygen during the chemical and structural transformations. Excellent quantitative agreement is now found between theoretical and measured values of coercivity, shape anisotropy alone seeming to be involved. Moreover, the qualitative agreement between theoretical values of intrinsic induction and remanence and those derived from actual powder dispersions is much better than with the earlier treatment.

---

§ 1. Introduction

In an earlier paper (Osmond 1952, to be referred to as I), a tentative explanation was offered of the observed magnetic properties of the fine powders of magnetite and γ-ferric oxide which are often dispersed in a non-magnetic medium to form the active coatings of sound-recording tapes. It was found that, while the ratios of the measured coercivities of γ-ferric oxide powders of quasi-spherical grains to those of markedly acicular powders were qualitatively in agreement with theoretical values based on shape anisotropy of the particles (Stoner and Wohlfarth 1947, 1948, Néel 1947), the actual and relative values of the coercivities of the acicular γ-ferric oxide powders and of the similar magnetite powders from which they were produced by oxidation were considerably lower than these theories would predict. It was therefore concluded that magnetocrystalline anisotropy played an important part, particularly in the case of γ-ferric oxide.

The vagueness of this last concept was unsatisfactory, and subsequent study of a recent review (Ervin 1952) of the transformations taking place in the dehydration of α-AlO.OH to α-Al_2O_3, which must equally apply to the isomorphous iron oxides (the synthetic forms of which are the origins of the acicular magnetic oxide powders studied here), has revealed a fundamental weakness of the previous treatment. It was tacitly assumed in I that each particle, of proved single ferromagnetic domain dimensions, behaved like a homogeneous single crystal, whereas it now appears that the structure probably becomes quasi-porous during dehydration and subsequent reduction, on account of the loss of oxygen.

The study of the properties of the acicular powders has therefore been undertaken again from this new point of view, with much improved results. As in I, the term 'coercivity' is used throughout this paper to denote the intrinsic coercive force (${}_IH_c$) when the previously applied magnetizing field has been three or more times as large as this amount; all values of field and coercivity are measured in oersteds and those of magnetic intensity and intrinsic induction in gauss.

§2. FORMATION OF CAVITIES

(i) *Changes of Structure*

Ervin (1952) has carried out a new detailed study by X-ray and optical analysis of the transformation by dehydration of diaspore (α-AlO.OH) to corundum (α-Al_2O_3), reaching the same conclusions as those found earlier by Deflandre (1932). It is well known (cf. Wells 1945) that these oxides of aluminium are isomorphous with the corresponding oxides of iron; hence the same changes should occur in the dehydration to synthetic α-Fe_2O_3 of the pigment oxides (synthetic α-FeO.OH) often used as the primary material for the production of acicular magnetic oxide powders. The basic structure, consisting of layers of close-packed oxygen atoms in hexagonal stacking, is shown to remain unchanged, and Ervin attributes the fact that the dehydrated oxide particles are pseudomorphs of the original monohydrate grains, which was noted in I on the evidence of electron microphotographs, to the circumstance that the transformation takes place *in situ*, so that major structural changes become impossible. Thus the *a* axis of diaspore, perpendicular to the basic oxygen layers (along which the synthetic monohydrate must initially have grown with the deposition of more oxygen layers), becomes the *c* axis of corundum, the axis of threefold symmetry. It must also follow, on the same reasoning, that the subsequent reduction to magnetite of the α-ferric oxide produced in the same way will lead to particles of similar close-packed layers of oxygen, but now in cubic stacking, the *c* axis of the rhombohedral structure becoming the [111] cubic solid diagonal, which is the direction of easy magnetization both in magnetite and in the γ-ferric oxide produced from it by oxidation without change of basic structure. Hence from synthetic monohydrate crystals elongated along the axis at right angles to the close-packed oxygen layers there are eventually derived magnetic oxide particles elongated along the directions of easy magnetization, a circumstance which justifies the assumption in I that these single domain acicular particles should behave like prolate rather than oblate spheroids, and consequently should exhibit pronounced coercivities due to shape anisotropy (Stoner and Wohlfarth 1948).

However, Ervin also finds that, although under X-rays the dehydrated particles behave like single crystals of corundum, yet optically they behave like two sets of crystallites with their twofold axes directed in opposite senses. Now the twofold rhombohedral axes are at right angles to the threefold *c* axis, so that this subdivision of the particle into smaller crystallites will not affect the uniformity of direction of the *c* axes which, as already stated, transform on reduction to the [111] cubic directions in magnetite. This subdivision is shown to be real from the mosaic pattern obtained with polarized light under the microscope with partially dehydrated diaspore, some crystallites being converted and some not. It follows that each acicular iron-oxide particle must be considered to consist of a collection of crystallites, all having their [111] cubic directions of easy magnetization aligned in the same sense, most probably parallel to the long dimension of the particle. Owing to the cohesion of the particle as a whole this collection of crystallites will behave magnetically like one single-domain ferromagnetic body, but the subdivision most probably entails small spaces between individual crystallites, caused initially by the loss of oxygen during the chemical changes, and these spaces will act as 'cavities' in an otherwise solid ferromagnetic particle which is too small for the spaces to be Bloch boundaries.

(ii) *Estimated Volume Concentration of Cavities*

An idea of the order of magnitude to be expected for the volume concentration of cavities in the various structures can be obtained by considering the sequence of chemical changes expressed by the following formulae:

(1) Dehydration: $6\alpha\text{-FeO.OH} \rightarrow 3\alpha\text{-Fe}_2\text{O}_3 + 3\text{H}_2\text{O}$.

(2) Reduction: $3\alpha\text{-Fe}_2\text{O}_3 + \text{H}_2 \rightarrow 2\text{Fe}_3\text{O}_4 + \text{H}_2\text{O}$.

(3) Re-oxidation: $2\text{Fe}_3\text{O}_4 + \text{O} \rightarrow 3\gamma\text{-Fe}_2\text{O}_3$.

Neglecting the relatively small hydrogen atoms in the monohydrate, one-quarter of the oxygen volume is lost during dehydration and one-ninth (one-twelfth of the original) during reduction, the latter being recovered during re-oxidation. Considering oxygen atoms alone, the total loss of volume in the formation of magnetite would be one-third, which is reduced to one-quarter on re-oxidation to γ-ferric oxide. However, it seems more accurate to take account of the volume of iron present in the structure which remains unchanged throughout the transformation. Since the ionic radius of Fe (whether divalent or trivalent) is approximately half that of the O^{2-} ion, the volume of the latter may be taken to be eight times that of an iron atom. Hence an initial volume equal to that of $12\frac{3}{4}$ oxygen atoms is reduced to $8\frac{3}{4}$ (a loss of 31·4%) in the transformation to magnetite, but this is restored to $9\frac{3}{4}$ (a total reduction of 23·5%) on re-oxidation to γ-ferric oxide.

Since the loss or gain of oxygen must take place at the surface as well as in the interior of the particle, it does not necessarily follow that the volume concentration of cavities in the structure is the same as the fractional volume lost during the transformations, but since the particles retain the same general shape throughout, the volume of cavities in the two magnetic oxides formed in the same cycle of changes may well be approximately the same fraction of the total loss of volume. Initially it will be assumed that this is so.

§ 3. THE EFFECT ON COERCIVITY

Weil (1951) has shown, with particular reference to the porous particles of precipitated cobalt produced by McCartney and Anderson (1947), that it is the internal cavities in such structures that have the most effect upon coercivity, and that the 'apparent density' formula due to Néel (1947) should apply to these, rather than to the spaces between individual particles. It was the volume concentration of the latter in typical tape coatings that was studied in I, and evidence for their relative unimportance is provided by the fact that the measured coercivities of powders of different tightnesses of packing in tubes are so little different from each other and from those of similar powders dispersed in tape coatings.

When the coercivity depends entirely upon shape anisotropy, the theoretical formula for an assemblage of identical single-domain particles containing a volume fraction V of cavities can be written

$$ {}_IH_c = pVN_dI_0, \qquad \ldots\ldots(1) $$

where I_0 is the saturation magnetization of the ferromagnetic substance excluding cavities, N_d is the difference of the demagnetizing factors along the equatorial and polar axes of the prolate ellipsoid to which each particle may be considered to approximate, and p is a factor depending upon the orientation of each particle relative to the field which it experiences. A single homogeneous particle can be considered as one of an assemblage of infinite dilution, i.e. $V=1$ in the above formula. Hence the effect of the cavities is to cause an apparent reduction of the saturation intensity of the material to an effective value VI_0.

For random orientation $p=0{\cdot}48$ (Néel 1947, Stoner and Wohlfarth 1948), while the values of I_0 for magnetite and γ-ferric oxide, calculated in I, are respectively 480 and 400. Assuming that in both oxides produced during one cycle of transformations the same fraction n of the total reduction in volume appears as cavities, while the mean shape factor remains the same, we find for the two theoretical coercivities:

$$\left.\begin{aligned}\text{Magnetite: } {}_IH_c&=0{\cdot}48N_d\times(0{\cdot}314\,n\,.\,480),\\ \gamma\text{-ferric oxide: } {}_IH_c&=0{\cdot}48N_d\times(0{\cdot}235\,n\,.\,400).\end{aligned}\right\}\quad\ldots\ldots(2)$$

The ratio of the latter to the former is 0·624, and this is *precisely the average value found in* I *for actual oxides*, after making allowance for departures from stoichiometric composition, whereas on the supposition of non-porous structure the ratio was expected to be 400/480 = 0·83. The small variations found in many experimental examples, ranging from 0·61 to 0·64 and usually lying between 0·62 and 0·63, are well within probable slight differences in the value of the fraction n for the two oxides.

Thus the new theory is entirely adequate to account for the observed ratio of coercivities of these two oxides. To examine actual individual coercivities in the light of this theory we will consider a pair of values studied in I. When corrected for non-stoichiometric composition these were 342·9 for magnetite and 214·4 for γ-ferric oxide, the ratio being 0·625.

Assuming first that $n=1$ (the total loss of volume appears as cavities), the values of N_d given by substituting these values in eqns. (2) are 4·741 for magnetite and 4·752 for γ-ferric oxide. As expected, these are virtually identical, and from the tables given by Stoner (1945) they correspond to a shape factor of approximately 3·75, which is about half the observed mean value for the particles themselves. It is of course possible that the average shape factor of the crystallites is thus reduced, but it seems equally likely that $n<1$, which entails a larger value of N_d. For instance, if the latter were 5·75, corresponding to the mean shape factor observed for usual powders (approximately 8), we should find $n\simeq 0{\cdot}825$, which seems a very reasonable value. On the other hand, for $N_d=2\pi$, the limiting value for long thin particles, $n=0{\cdot}756$, which must be the lower limiting value if the present theory holds.

If the volume concentration of spaces between oxide particles in the tape coating had to be imposed upon that of the cavities in the particles, the resulting values of N_d would exceed 2π even when $n=1$, using the values calculated in I, and they would be still larger using the new values of oxide density resulting from the porous structure. This provides further evidence for the relative unimportance of the powder packing; in other words the demagnetizing effect of one particle upon another appears negligible.

As a further illustration it was noted in I that the coercivities of tapes of the original German type, coated with almost spherical particles of γ-ferric oxide produced from directly precipitated magnetite, are of the order of 90 compared with an average value of 225 for tapes coated with acicular particles. Assuming similar departures from stoichiometric composition, and that the directly precipitated particles are non-porous, so that $VI_0 = 400$, we find for the German oxide $N_d \simeq 0{\cdot}4 \times 0{\cdot}235 \times 4{\cdot}75$ (or the equivalent $0{\cdot}4 \times 0{\cdot}825 \times 0{\cdot}235 \times 5{\cdot}75$), i.e. 0·446, which corresponds to a shape factor of rather less than 1·1. On the old assumption of equally non-porous acicular particles we found $N_d = 0{\cdot}4 \times 5{\cdot}75 = 2{\cdot}30$, for which the shape factor is approximately 1·65. The lower value seems much more likely.

Hence the new theory fits both actual and relative coercivities much better than the old treatment, and from the point of view of shape anisotropy alone.

§4. The Effect on Magnetization Curves for Coated Tapes

In I the theoretical values of intrinsic induction ($B_i = 4\pi I$) at magnetizing fields of 1000 and at remanence (B_r) were calculated for tapes, for comparison with those given by integrated cathode-ray tube traces showing the variation of dI/dt with alternating fields for two particular tapes one of which was coated with magnetite (giving a known value of remanent flux) and the other with γ-ferric oxide. It was assumed that the thickness of each tape coating was 0·7 mil (1 mil $= 10^{-3}$ in.), a fairly usual value, and the resulting values of B_i and B_r were noticeably lower than the theoretical figures. For future comparison with theoretical values calculated by both old and new treatments these actual values are now repeated for three possible tape thicknesses.

Thickness (mil)	Magnetite tape B_r	Magnetite tape B_i	γ-ferric oxide tape B_r	γ-ferric oxide tape B_i
0·7	550	1092	565	1125
0·65	592	1176	608	1212
0·6	642	1274	659	1313
	$B_r/B_i = 0{\cdot}504$		$B_r/B_i = 0{\cdot}502$	

In the old treatment it was considered that account should be taken of the probable interactions between particles instead of adopting the simple random orientation approach to saturation given by Stoner and Wohlfarth (1948), while still retaining their theoretical value of 0·5 for the ratio of remanence to true saturation. For this purpose the law of approach to saturation (taking account of cavities) due to Néel (1948) was used with the following results. These are the values calculated in I extended to include stoichiometric and two cases of non-stoichiometric composition for each oxide. The figures shown for B_r are half the calculated saturation values.

Magnetite tape

FeO content (%)	B_r	B_i	B_r/B_i
31	775	1321	0·586
29·5	721	1247	0·578
29	704	1221	0·576

γ-ferric oxide tape

FeO content (%)	B_r	B_i	B_r/B_i
0	667	1173	0·569
1	691	1212	0·570
2	715	1248	0·573

On the new theory the interactions between particles will be as unimportant during the approach to saturation as they apparently are in their effect on coercivity. However, the volume concentration of oxide particles in the tape coating, now denoted by $1 - U$ to distinguish it from the concentration $1 - V$ of magnetic substance in each particle, must be used to estimate the intrinsic induction of the *tape* after calculating that of a single particle. This volume concentration of particles will of course change with every change of volume of cavities in each particle, owing to the resultant change in density.

Néel's formula was developed to describe the approach to saturation by rotational processes in polycrystalline materials, and good agreement with it was found in experiments on samples of sponge iron of different degrees of porosity. In the present study each oxide particle is considered to consist of a number of crystallites, and may thus be compared to a very small portion of such porous iron. The formula thus seems ideal for estimating the magnetization of an individual particle, more particularly at fields well above the coercivity value. In the form given by Stoner (1950) it is

$$1-\frac{I}{I_m}=\frac{V}{2(1-V)}\left[\frac{2+3\alpha}{4(1+\alpha)^{1/2}}\ln\left\{\frac{(1+\alpha)^{1/2}+1}{(1+\alpha)^{1/2}-1}\right\}-\frac{3}{2}\right], \qquad \ldots\ldots(3)$$

where I_m is the saturation magnetization of the particle, i.e $(1-V)I_0$, and $\alpha=H/4\pi I_m$. In what follows we will refer to the fractional magnetization I in this formula as the 'Néel' magnetization, with the symbol I_N, and will denote by I_P its value at the initial peak magnetizing field. We can now apply a modified form of Stoner and Wohlfarth's treatment, the omission of which was a further weakness in I. As the field is reduced from its initial positive maximum value ($H=1175$ in I) the absolute value of the magnetization vector remains at I_P instead of I_0 until the reduced field $h=H/VN_dI_0$ reaches a value close to $-0{\cdot}48$ (the energy considerations giving the actual critical value in this case are beyond the scope of this present study). When the discontinuous jump in the resolved magnetization has taken place, the new value will no longer be proportional to I_P but to the value of I_N appropriate to the value of H at that instant, and this will increase as the magnitude of H increases. The result will be to open the tip of the hysteresis loop from the coincidence shown for $|h|\geqslant 0{\cdot}9$ in Stoner and Wohlfarth's curve, since I_N will be numerically less than I_P until $|H|$ reaches its initial peak value. This is, of course, what is seen in practice, although much of the observed difference in ordinates is doubtless due to variations in shape factor, exact composition, etc., between individual particles.

Having calculated I_P, and I_N at $H=1000$, the mean of these values, multiplied by the Stoner and Wohlfarth value of $\cos\phi$ for the appropriate value of h and by $4\pi(1-U)$, is the theoretical value of B_1 for the tape corresponding to those measured in I. Similarly $B_r=0{\cdot}5\,(1-U)4\pi I_P$. Since each change in V involves a change in U, calculation becomes somewhat tedious for a range of values of V combined with varying degrees of departure from stoichiometric composition. The main details for one combination are given as an illustration, the results of six different calculations for each oxide being shown in the tables which follow.

Example. Magnetite tape with 29·5% FeO content, $V=0{\cdot}259$ (i.e. $n=0{\cdot}825$), weight ratio of oxide to binder $=2$, ${}_IH_c=325$.

In I we calculated

$I_0=0{\cdot}93\times 480=447$; hence $I_m=(1-V)I_0=0{\cdot}741\times 447=331$; $4\pi I_m=4158$.

At $H=1175$, $\alpha=0{\cdot}283$; hence $I_P=0{\cdot}956\times 331=316{\cdot}5$.

At $H=1000$, $\alpha=0{\cdot}241$; hence $I_N=0{\cdot}950\times 331=314{\cdot}5$.

Also, since ${}_IH_c=325$, for which $h=0{\cdot}48$, it follows that at $H=1000$, $h=1{\cdot}477$, for which the Stoner and Wohlfarth value of $\cos\phi$ is 0·963.

Again, in I the density of magnetite was calculated as 5·206; hence the density of a particle in this case is $0{\cdot}741 \times 5{\cdot}206 = 3{\cdot}858$. Assuming a density of 0·9 for the binder materials, this leads to a volume concentration of particles in the tape of $1 - U = 0{\cdot}3182$.

Hence at $H = 1000$, $B_i = 0{\cdot}3182 \times 0{\cdot}963 \times 4\pi \times 315{\cdot}5 = 1214$,
and $B_r = 0{\cdot}5 \times 0{\cdot}3182 \times 4\pi \times 316{\cdot}5 = 633$; $B_r/B_i = 0{\cdot}521$.

Magnetite tape

	$V=0{\cdot}314$ ($n=1$)			$V=0{\cdot}259$ ($n=0{\cdot}825$)		
FeO content (%)	B_r	B_i	B_r/B_i	B_r	B_i	B_r/B_i
31	654	1255	0·5213	678	1301	0·5208
29·5	611	1170	0·5216	633	1214	0·5210
29	597	1145	0·5219	619	1187	0·5212

γ-ferric oxide tape

	$V=0{\cdot}235$ ($n=1$)			$V=0{\cdot}194$ ($n=0{\cdot}825$)		
FeO content (%)	B_r	B_i	B_r/B_i	B_r	B_i	B_r/B_i
0	595	1164	0·5107	609	1193	0·5101
1	615	1205	0·5109	630	1234	0·5102
2	636	1246	0·5107	652	1276	0·5105

Comparison of all these results shows that the values derived from the new theory of porous particles are much closer to those calculated from actual tapes than are the figures given by the older theory, in particular the relative values of B_r and of the ratio B_r/B_i. For the measured tapes the remanence of the γ-ferric oxide coating was 1·027 times that of the magnetite coating, and this ratio is given almost exactly by the new values for 2% FeO with $n = 1$ and 29% FeO with $n = 0{\cdot}825$, and also by those for 2% FeO and 29·5% FeO, both with $n = 0{\cdot}825$, no other pairs of values being so good. From remanence values, therefore, the actual tapes examined fit into the new picture as consisting of slightly non-stoichiometric oxides in coating thicknesses lying between 0·6 and 0·65 mil.

The measured values of B_r/B_i remain a little lower than any obtained theoretically. This may be a measuring fault due to errors in integration or to a slight lack of balance between the compensating and search coils of the magnetizing assembly, leading to a small tilt in the cathode-ray oscilloscope trace. However, it may be a physical reality due to a slight degree of preferred orientation of the particles or to the presence of a small proportion of ultra-fine particles (Néel 1949 a, b). Such small particles, behaving as paramagnetics, will make no contribution to the remanence of the powder as a whole but a small contribution to the intrinsic induction at higher fields. Hence, compared with a completely ferromagnetic powder, the ratio B_r/B_i will be lowered, B_r will be lowered by an amount roughly proportional to the volume fraction of fine particles, while ${}_IH_c$ will also be slightly reduced (Weil 1951).

§ 5. CONCLUSION

The new theory advanced in this paper, that the acicular powders of the magnetic oxides of iron produced from dehydrated synthetic goethite have a porous structure, containing cavities caused by the loss of oxygen during the various chemical processes, gives excellent quantitative agreement between theoretical and measured values of coercivity. Particle shape now appears to be the only important form of anisotropy in this connection, and the mutual interactions between particles dispersed in tape coatings of normal concentration seem to be relatively unimportant.

The agreement between values of remanence and of intrinsic induction at magnetizing fields of 1000 oersteds based on this new theory and those found for actual tapes coated with these oxides, although not so close as that between coercivities, is far better than that found on the older theory of solid particles with mutual interactions.

The basic interpretation of the magnetic properties of these dispersed powders remains unaltered, namely that the latter provide examples of a material consisting of single domain ferromagnetic particles embedded in a non-magnetic matrix. The new viewpoint is that cavities in the particles are far more important than the interparticle spacing in the medium.

ACKNOWLEDGMENTS

I am grateful to my colleague Mr. N. T. Winkworth for first bringing to my notice the possibility of this porous structure arising from the chemical transformations. This paper is published by permission of the Directors of the Minnesota Mining and Manufacturing Co., Ltd.

REFERENCES

DEFLANDRE, M., 1932, *Bull. Soc. Franç. Min.*, **55,** 140.
ERVIN, G., 1952, *Acta Crystallogr.*, **5,** 103.
MCCARTNEY, J. T., and ANDERSON, R. B., 1947, *J. Appl. Phys.*, **18,** 902.
NÉEL, L., 1947, *C. R. Acad. Sci., Paris*, **224,** 1550; 1948, *J. Phys. Radium*, **9,** 184; 1949 a, *C. R. Acad. Sci., Paris*, **228,** 664; 1949 b, *Ann. Géophys.*, **5,** 99.
OSMOND, W. P., 1952, *Proc. Phys. Soc.* B, **65,** 121.
STONER, E. C., 1945, *Phil. Mag.*, **36,** 803; 1950, *Rep. Prog. Phys.*, **13,** 83 (London: Physical Society).
STONER, E. C., and WOHLFARTH, E. P., 1947, *Nature, Lond.*, **160,** 650; 1948, *Phil. Trans. Roy. Soc.* A, **240,** 599.
WEIL, L., 1951, *J. Phys. Radium*, **12,** 437.
WELLS, A. F., 1945, *Structural Inorganic Chemistry* (Oxford: Clarendon Press), pp. 332, 356.

Low Temperature Conduction in Extremely Degenerate Semiconductors

By R. W. WRIGHT
University College, Ibadan, Nigeria

MS. received 29th August 1952

Abstract. The variation of conductivity of cadmium oxide with temperature has been measured over the temperature range 100°–800°K. It is found that the variations observed in the specimens used can be accurately predicted if it is assumed that all the electrons which are effective in the conduction process have energies above the bottom of the conduction band exceeding the characteristic energy of the lattice $h\nu$. This appears to be the first test of the mean free path theory for such electrons in ionic lattices. The properties of these electrons are markedly different from those with energies less than $h\nu$ which are usually met with in semiconductors.

§1. Introduction

The values of the conductivity predicted by the Lorentz–Sommerfeld theory for electrons in ionic latticed materials depend upon the mean free path theory assumed for the electrons. It is therefore of importance to test the agreement between theoretical and experimental values of the conductivity because this is also a verification of the mean free path theory used. The mean free path theory which an electron in an ionic lattice obeys depends upon whether the energy E, which it possesses above the bottom of the conduction band, is greater or less than $h\nu$. ν is the characteristic frequency of the lattice. The first theory which assumes that $E<h\nu$ was shown by Mott and Fröhlich (1939) to be in good agreement with the experimental results obtained for Cu_2O by Englehard (1933). This theory is applicable to the majority of ionic latticed electronic conducting materials, since the density of free electrons is usually very small and therefore E is generally less than $h\nu$. However, so far as the author knows, no comparison has been made between the results of experiment and those predicted by the theory which assumes that $E>h\nu$. Although it is obvious that E cannot be greater than $h\nu$ for all electrons, it is possible that this may be so for all the electrons which are effective in the conduction process. It will be shown that the electrons in cadmium oxide are such that if the second theory, where $E>h\nu$, is assumed for all free electrons a variation of conductivity with temperature is predicted which is closely followed by the experimental results obtained.

§2. Theory

The mechanism of conduction in CdO in the form of sintered pressed powder is very sensitive to slight differences in the mode of preparation. Specimens prepared by various investigators appear to show a gradation of properties. Blakemore (1951), who used very little sintering, found that his specimens were fairly typical semiconductors with log σ proportional to $1/T$, but at high temperatures there were slight indications that the conductivity failed to increase with sufficient rapidity to obey this law. Hogarth (1951), who sintered his specimens more heavily, found that they behaved normally at low temperatures, but as the temperature increased the conductivity passed through a maximum. The author (1951 a), who was particularly concerned with very heavily sintered specimens, has found that in such specimens over a very wide temperature range (100°–800°K) the conductivity decreases with increasing temperature. This paper is concerned with such heavily sintered specimens and it is shown that the variation with temperature is in complete accordance with the second theory of mean free path of electrons in ionic lattices mentioned above.

In these well sintered specimens, the author has shown from measurements of the Hall coefficient that the concentration of free electrons is constant over a range of temperatures 290°–800°K, and furthermore, from the values of the conductivity and thermoelectric power over the same range, the free electrons are extremely degenerate. The mechanism that has so far best explained these results is to consider that the impurity centres (excess Cd atoms) are packed so closely together in the lattice that the valency electrons are always free, so that the specimen is like an 'impurity metal'. This mechanism is similar to the one proposed by Pearson and Bardeen (1949) for the more impure specimens of silicon. Cadmium oxide has an ionic lattice, which complicates further consideration of the properties because, as mentioned above, the mean free path theory depends upon whether the electron has an energy E greater or less than $\boldsymbol{h}\nu$. The expressions for the mean free path of electrons in ionic lattices has been worked out for both cases by Mott and Fröhlich (1939), and Davydov and Shmushkevitch (1940). By assuming that the vast majority of the electrons which effectively contribute to the electrical properties have energies E greater than $\boldsymbol{h}\nu$, and using the appropriate mean free path formula in the Lorentz–Sommerfeld theory, the author has shown that, over a temperature range in which a certain approximation may be made, the resulting expressions for the electrical properties agree very well with those observed by experiment. Also if it is assumed that all the electrons have energies less than $\boldsymbol{h}\nu$ it is impossible to explain the experimental values obtained for the thermoelectric power since they are much lower than any of the values predicted by the theory. Thus, we may assume that not only is there a constant number of free electrons, but these electrons are very degenerately distributed in their energy band, with the level of the Fermi energy E^* greater than $\boldsymbol{h}\nu$. Now in a degenerate electron gas, only electrons with energies in the neighbourhood of E^* are effective in determining the electrical properties, and here the effective electrons will have E greater than $\boldsymbol{h}\nu$. This, however, will be more accurately true at low temperatures when the electrons are more highly degenerate.

The expression for the mean free path on the assumption that $E > \boldsymbol{h}\nu$, is:

$$l(E)=l_0\eta(2T/\Theta)\{1+2/[\exp(\Theta/T)-1]\}^{-1} \qquad \ldots\ldots(1)$$

where l_0 is a constant depending upon the lattice structure, $\eta=E/\boldsymbol{k}T$ the 'reduced' energy, and $\Theta=\boldsymbol{h}\nu/\boldsymbol{k}$ a characteristic of the lattice.

Assuming that the density of free electrons is constant and that they are very degenerate, the Lorentz–Sommerfeld theory leads to

$$\rho=\rho_0\Theta[\tfrac{1}{2}+\{\exp(\Theta/T)-1\}^{-1}]=\rho_0\Theta\beta \qquad \ldots\ldots(2)$$

as an expression for the resistivity. When $T \gg \Theta$ this approximates to

$$\rho=\rho_0 T \qquad \ldots\ldots(3)$$

and even when $T=\Theta$, $\rho=1{\cdot}082\rho_0 T$.

Thus, for temperatures T greater than Θ, the theory predicts that eqn. (3) is a good approximation. This was shown by the author (1951 a) to be in reasonable accord with experiment over the temperature range 290°–800°K. It is the object of this paper to consider the more accurate expression (2) in relation to experimental results over a wider temperature range 100°–800°K, over parts of which eqn. (3) is not even a rough approximation.

A further factor which must be taken into account is the presence of the impurity centres themselves. We may use the simple formula as used in the theory of alloys and assume that this impurity scattering will have a constant effect ρ^* upon the resistivity, thus

$$\rho = \rho^* + \rho_0 \Theta \beta. \qquad \qquad \ldots\ldots (4)$$

This equation is quite consistent with the argument of Jones (1951) as it is a special case in which the energy of the effective electrons, and hence the mean free paths due to both lattice and impurity scattering are very nearly independent of temperature. To compare eqn. (4) with experiment, graphs of ρ against β are plotted for different values of Θ.

§ 3. EXPERIMENTAL RESULTS AND DISCUSSIONS

The conductivity was measured in the usual manner, using potential probes. The specimens were heavily sintered and were therefore so coherent that they could be filed into exact shapes.

Figure 1 shows experimental results for the variation with temperature of the resistivity of specimen I over the temperature range 100°–750°K compared with the thick line which represents the variation predicted by theory assuming the constants shown in the table. This was the only specimen investigated over the whole range, as the resistivity at temperatures above room temperature had been investigated previously (Wright 1951 a). It will be seen that at temperatures above room temperatures it is a fair approximation to say that the resistivity is proportional to the temperature, as predicted by eqn. (3). Figure 2 shows the variation of resistivity for Specimens I and II over the temperature range 100°–300°K together with the variations predicted theoretically, assuming the values in the table. Figures 3 and 4 show the corresponding graphs of resistivity plotted against β for the data of the previous figures. The value of Θ chosen is that which gives the best straight line for the results. It will be seen that eqn. (4) is followed very accurately.

The values of the constants are shown in the table.

Specimen	Θ	ρ_0	ρ^*
I (run a)	520	$6{\cdot}39\times10^{-6}$ ohm cm	$-3{\cdot}52\times10^{-4}$
I (run b)	520	$6{\cdot}39\times10^{-6}$ ohm cm	$-3{\cdot}42\times10^{-4}$
II (run a)	430	$1{\cdot}70\times10^{-6}$ ohm cm	$-0{\cdot}80\times10^{-4}$

The values of ρ^* are of interest in that they have negative values. In a previous letter (Wright 1951 b) the author deduced positive values of ρ^* using data obtained at temperatures above room temperature, assuming that $T \gg \Theta$. It was not at the time realized that Θ was so large, and thus although the graphs are surprisingly linear, the value of Θ affects the intercept. The deduction of ρ^* from the results is therefore not so simple as was indicated. The question of the origin of these negative values of ρ^* is not very obvious. There appear to be three possibilities: (i) ρ^* may not be a constant since it is assumed that the cross section of Rutherford scattering for electrons of constant energy is independent of temperature; the impurity will have thermal energy and thus its vibrations will be temperature dependent, which will affect its effective cross-sectional area. (ii) The mean free path theory used is possibly inaccurate at temperatures of the order of Θ and greater in this region, since it may predict values of the mean free path of the order of the interatomic distance, where the

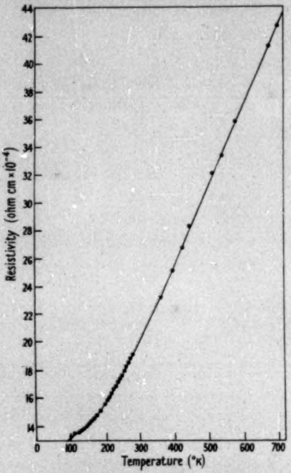


Fig. 1. Graph of resistivity against absolute temperature for Specimen I, showing experimental points in relation to the curve predicted by theory.

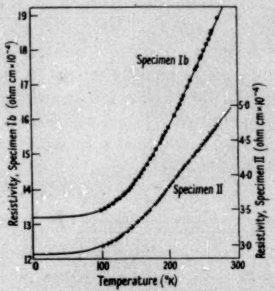


Fig. 2. Graph showing experimental values of resistivity against absolute temperature of Specimens Ib and II in relation to the curves predicted by theory assuming the values of the constants given in the table.

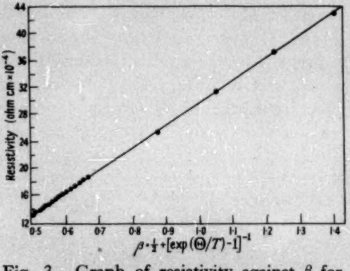


Fig. 3. Graph of resistivity against β for Specimen I showing points derived from experimental results.

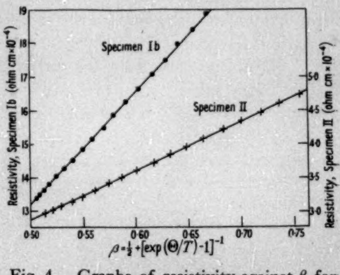


Fig. 4. Graphs of resistivity against β for Specimens Ib and II showing points derived from experimental results.

approximations made in the deduction of the mean free path theory are no longer valid. (iii) The cause may be the presence of the impurity centres themselves, after the valency electrons have been liberated. It is usually considered that their presence will increase the resistivity due to the fact that they break up the perfect periodicity of the lattice. However, in the case of the specimens of CdO considered here, the quantity of 'impurity' or excess cadmium is very great, since (*a*) the free electrons are always free, (*b*) they are degenerate, and (*c*) $E^* > \boldsymbol{h}\nu$. It is therefore possible that presence of such a large quantity of impurity centres may have other effects which would decrease the resistivity, such as increasing the density of energy levels available to the electrons, or altering the lattice structure, so that in some way it is possible that the negative value of ρ^* may be explained.

It will be noticed that one of the specimens has an extremely high conductivity, about 2000 ohm^{-1} cm^{-1} at room temperature and increasing to about 3000 ohm^{-1} cm^{-1} at 100°K. Specimens of CdO having conductivities nearly as high have been prepared by Preston (1950) by a sputtering process, and therefore these values are not abnormal or peculiar to the process of making. Preston's specimens were thin transparent films of CdO but were, however, unstable when heated. Similar results have been obtained by Helwig (1952) who has investigated the irreversible sections of the heating. These results lead one to the conclusion that many of the intermediate compounds such as some of the CdO specimens, and possibly such materials as MnO_2, PbO_2, etc., will form a special type of semiconductor in which n_b, the number of impurity centres, will vary with temperature as the lattice changes from one form of oxide to another. This variation could be reversible or irreversible and would be dependent on the variation of the strength of the chemical bond with temperature which might thus be investigated electrically.

From the results given here, the variation with temperature of the expression given in eqn. (1) for the mean free path derived by Mott and Fröhlich and Davydov and Shmushkevitch for electrons in ionic crystals with energies E greater than $\boldsymbol{h}\nu$, may be considered in agreement with that observed experimentally.

ACKNOWLEDGMENTS

The major experimental portion of this work was carried out in the Physics laboratories of Queen Mary College, London, by kind permission of Professor H. R. Robinson, to whom I am greatly indebted for both facilities and encouragement. I would also like to thank Professor J. P. Andrews and Drs. Blakemore, Hogarth and Putley for very helpful discussions.

REFERENCES

BLAKEMORE, J., 1951, *Thesis*, University of London.
DAVYDOV, B., and SHMUSHKEVITCH, I., 1940, *J. Phys., USSR*, **3**, 359.
ENGELHARD, E., 1933, *Ann. Phys., Lpz.*, **17**, 501.
HELWIG, G., 1952, *Z. Phys.*, **132**, 621.
HOGARTH, C. A., 1951, *Nature, Lond.*, **167**, 521.
JONES, H., 1951, *Phys. Rev.*, **81**, 149.
MOTT, N. F., and FRÖHLICH, H., 1939, *Proc. Roy. Soc.* A, **171**, 496.
PEARSON, G. L., and BARDEEN, J., 1949, *Phys. Rev.*, **75**, 865.
PRESTON, J. S., 1950, *Proc. Roy. Soc.* A, **202**, 449.
WRIGHT, R. W., 1951 a, *Proc. Phys. Soc.* A, **64**, 350; 1951 b, *Ibid.*, **64**, 949.

The Origin of Secondary Emission Electrons

By A. LEMPICKI
Electronic Tubes Ltd., High Wycombe, Bucks.

MS. received 4th December 1952

Abstract. A secondary electron multiplier valve is used to determine the energy necessary to liberate secondary electrons from emitting surfaces. It is found that in the case of BaO and MgO secondary electrons originate from the valence band and not from impurity centres.

The great secondary emission yields obtained from insulators have been ascribed by Bruining and de Boer (1939) to a reduced interaction between the internal secondaries and the electrons in the conduction band. In good insulators the latter has a very small electronic population and the secondaries stand a better chance of escape before they lose energy. The rule laid down by these authors is that a substance will have a large secondary emission ratio δ if the long wavelength limit of the photoelectric effect coincides with the optical absorption limit. If, on the other hand, the absorption begins at longer wavelengths than the photoelectric effect, the substance will have a low δ.

The rule of Bruining and de Boer presupposes that the secondary electrons originate from the valence band of the substance and not from impurity centres.

A simple experiment to test this hypothesis can be performed by using a secondary emission valve similar to that designed by Bull and Atherton (1950), and having a secondary emitting surface made of a mixture of BaO and MgO. The circuit is given in fig. 1. Keeping the potential difference between the secondary emitter and anode constant (100 v) the current flowing in the secondary emitter lead was observed while varying the primary electron energy, as determined by the voltage V_p. We may expect a sudden rise in emission, i.e. in current i_{SE}

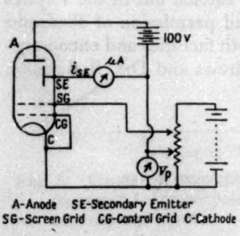

A
100 v
μA
i_{SE}
SE
SG
CG
C
V_p
A-Anode SE-Secondary Emitter
SG-Screen Grid CG-Control Grid C-Cathode

Fig. 1.

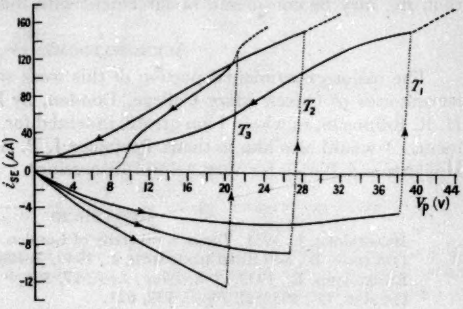


Fig. 2.

when the primary energy reaches a value equal to the energy necessary for the formation of secondaries. The magnitude of this energy should indicate whether the secondaries originate from the valence bands or from the higher impurity levels or electron traps. We may also expect that the direction of the current i_{SE} will change when δ becomes larger than unity.

The results of these measurements are given in fig. 2. Consider first the curve marked T_1. As we increase V_p the current i_{SE} slowly decreases from zero (electronic flow out of the secondary emitter) up to a point (in this case 38 v) when a sudden reversal in direction takes place and δ becomes larger than unity.

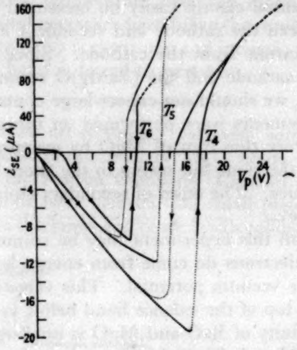


Fig. 3.

In this, as in the other curves, the dotted portions correspond to regions of instability where no accurate measurements could be made. If the voltage V_p is increased still further the current will rise, as indicated by the broken line. If, however, V_p is decreased it will fall without changing its direction. This indicates that δ still remains larger than unity. The kind of hysteresis loop which we obtain is an indication that V_p does *not*, in fact, correspond to the energy of the primary electrons. The charging of the surface of the secondary emitter is responsible for this effect. This can be demonstrated further by warming it up on passing current through a heater fitted inside the secondary emitter electrode. Curves T_2 and T_3 were obtained in this manner. As the temperature is increased the voltage which the coating may withstand (before the reversal of the current) and the area of the loop, gradually decrease. By continuing this process the curves T_4, T_5 and T_6 of fig. 3 were obtained, each of them corresponding to a higher temperature. Finally a temperature* is reached (T_6) when the loop disappears and the same curve can be retraced in both directions. We can assume that in this case V_p corresponds to the energy of the primary electrons and that the resistance is low enough to prevent the formation of surface charges. The sudden, but in this case reversible, rise in i_{SE} occurs at $V_p \simeq 10$ v and it represents the energy necessary to liberate secondaries from the target.

---

* It is estimated that T_6 corresponds to a temperature of about 300°c.

There are several remarks which can be made about this experiment:

(i) The use of small currents was deliberate in order to reduce, as far as possible, a variable heating of the target as i_{SE} changes with V_{p}.

(ii) All the curves except T_6 indicate that when the reversal point is reached a very large change in i_{SE} occurs (note the change of scale of i_{SE} above and below the zero line). This means that δ increases to values considerably larger than unity probably because of field enhanced secondary emission.

(iii) The value of 10 ev which we have obtained above should be corrected for both secondary electron emission energies and contact potential. The first are stated to be small for insulators (McKay 1948) of the order of a fraction of 1 ev. The contact potential cannot easily be measured directly because of the presence of grids between the cathode and secondary emitter which will tend to intercept all the electrons from the cathode. Since, however, the electron affinities for the oxide cathode and the (BaMg)O secondary emitter are likely to be very nearly equal, we should not expect large contact potentials.

(iv) Similar measurements were performed on tubes having the secondary emitter coated with a very thin film of MgO by evaporation in vacuum. The results were essentially the same but owing to the short life of these tubes were more difficult to reproduce. The onset of secondary emission appeared to occur in this case at about 11 ev.

The conclusions from this experiment may be summarized as follows: the bulk of the secondary electrons do come from energy levels which are situated at about 10 ev below the vacuum potential. This value is probably larger than the optical depth of the top of the valence band below vacuum potential. If we assume that electron affinity of BaO and MgO is not larger than 0·5 ev (Wright 1948) the top of the full band should be situated at about 6 ev in MgO (Weber 1951) and about 5 ev in BaO (Apker *et al.* 1951). The higher secondary emission threshold energy indicates that electrons are liberated from levels laying deeper than the top of the valence band. This may indicate the existence of some selection rules similar to those deduced by Wooldridge (1939) for metals.

ACKNOWLEDGMENT

The author is indebted to Electronic Tubes Ltd. for permission to publish these results.

REFERENCES

APKER, L., TAFT, E., and DICKEY, J., 1951, *Phys. Rev.*, **84,** 508.
BRUINING, H., and DE BOER, J. H., 1939, *Physica*, **6,** 834.
BULL, C. S., and ATHERTON, A. H., 1950, *Proc. Instn. Elect. Engrs.*, **97,** Pt. III, 65.
MCKAY, K., 1948, *Advances in Electronics*, I (New York : Academic Press), p. 105.
WEBER, H., 1951, *Z. Phys.*, **130,** 392.
WOOLDRIDGE, D. E., 1939, *Phys. Rev.*, **56,** 562.
WRIGHT, D. A., 1948, *Proc. Phys. Soc.*, **60,** 13.

The Electrical Conductivity of MgO Single Crystals at High Temperatures

By A. LEMPICKI
Electronic Tubes Ltd., High Wycombe, Bucks.

MS. received 4th December 1952

Abstract. The variation of electrical conductivity of MgO single crystals is measured as a function of temperature over the range from 300° to 1500°K. Heat pre-treatment is found necessary to stabilize the results at high temperatures. Assuming that the conductivity is electronic and intrinsic the energy gap between the full and conduction band is found to be 4·6 ev.

In connection with work on secondary cathodes some measurements were made of the electrical conductivity of MgO single crystals supplied to us by the courtesy of the Norton Company (Canada) in pieces ranging from 6 to 10 cm³. The crystals were cleaved into small plates of a few millimetres square and about 0·5 mm thickness and coated with colloidal graphite on opposite faces.

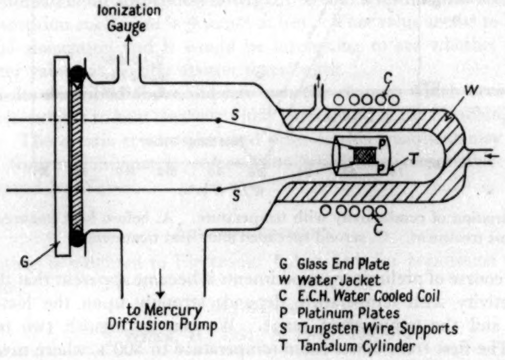


Fig. 1. Experimental arrangement used for measurement of MgO conductivity.

Figure 1 represents the arrangement by means of which all the measurements reported below were made. The crystal was held between two platinum plates inside a tantalum tube which was heated by eddy currents. The glass envelope was water cooled. The temperature was measured by means of a Pt, Pt–Rh thermocouple attached to one of the plates P (not shown on the drawing). The temperature of the crystal could be adjusted by varying the h.f. current in the coil C. The maximum crystal temperature attainable was about 1500°K. The system had the great advantage that this temperature could be maintained for hours without any part of the system becoming overheated. Positive ion emission from the tantalum tube disappeared after heating for about a day, and the electron emission, never very large, could be suppressed by maintaining a sufficiently high pressure in the system, of the order of $2\text{–}3 \times 10^{-4}$ mm Hg. A variable d.c. potential was applied to the plates and the current was measured by means of a Kipp galvanometer having a maximum sensitivity of 10^{-9} A cm^{-1}. The leakage currents were measured with the crystal removed over the whole temperature range. These proved to be substantially constant and mainly due to leakages outside the demountable system. On account of these leakages it was found inadvisable to use voltages across the crystal greater than 10 volts. Trials to check the linearity of current with voltage revealed that the crystals obey Ohm's law, at least up to fields of the order of 8000 v cm^{-1}. This constitutes sufficient reason to assume that contact resistances were absent.

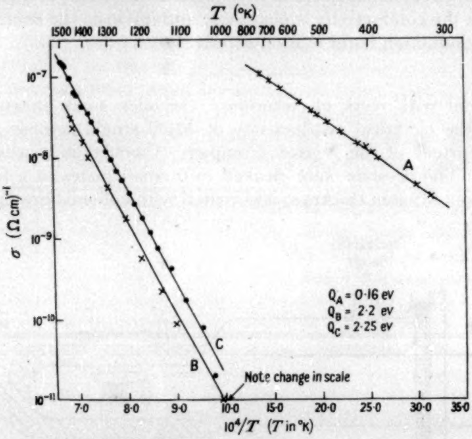


Fig. 2. Variation of conductivity with temperature. A, before heat treatment. B, after heat treatment. C, second specimen after heat treatment.

In the course of preliminary experiments it became apparent that the variation of conductivity with temperature depends strongly upon the history of the specimen and the temperature range. We can distinguish two temperature ranges. The first from about room temperature to 800°K where measurements are not very reliable and reproducible and the second from 1000°K to 1500°K where the converse is true.

One of the difficulties of measuring conductivity in the lower temperature range was due to the somewhat poor stability of the h.f. oscillator generating the eddy currents at small power output. The measurements were, therefore, carried out by bringing the crystal to a temperature of, say, 1000°K and then letting it cool and observing the current at the same time. Curve A of fig. 2 represents a typical result of this procedure. When the experiment is repeated we usually obtain a different value of the slope. Assuming an expression of the form $\sigma = \sigma_0 \exp(-Q/\boldsymbol{k}T)$ the values of Q fall between 0·15 and 0·25 ev.

In the higher temperature range the measurements were carried out by gradually increasing the temperature of the crystal from about 1000°K to 1400°K.

The results were, however, also very erratic. In view of this it was decided to subject the crystals to a heat treatment prior to the conductivity measurements. Thirty minutes of heating at approximately 1450°K seemed to stabilize the properties of the crystals to a very considerable extent. Curve B, fig. 2, was obtained after such heat treatment, using the same specimen as in the case A. The curve C gives results for a different crystal after similar heat treatment. The slopes of other curves all fall in the range 2·2–2·3 ev.

The stabilization of results obtained by heat treatment indicates that the original crystals deviate to some extent from the equilibrium composition at higher temperatures. The heating produces a more perfect crystal.

It is not possible to conclude from these experiments alone whether this high temperature conductivity is of ionic or electronic nature. If, however, we represent the conductivity of intrinsic semiconductors by the usual expression $\sigma=\sigma_0\exp(-\Delta E/2\boldsymbol{k}T)$ where σ_0 is a slowly varying function of temperature and ΔE the energy separation of the full and conduction band, we obtain $\Delta E \sim 4{\cdot}6$ ev. This value is, as we may expect, somewhat smaller than that deduced from optical measurements. Weber (1951), who investigated the optical absorption of MgO, found that bands due to excess of oxygen or magnesium are 'bleached' when the crystals are heated in vacuum at a temperature of 1500°K. One type of absorption which cannot be bleached has its long wavelength tail located between 5 and 4 ev and a steep rise commences at 6 ev. It is tempting to interpret this as the fundamental absorption of MgO. Weber's results indicate, however, that the absorption coefficient is 4 mm^{-1} at 6 ev. This value seems to be low for fundamental absorption and it would be interesting to see whether it rises to much greater values at slightly shorter wavelength.

Our results described above indicate that the changes which occur in MgO crystals when subject to heat treatment may correspond to the bleaching observed by Weber. The erratic results observed prior to this treatment may be due to departures from stoichiometric composition which are responsible for other bands observed by Weber.

ACKNOWLEDGMENT

The author is indebted to Electronic Tubes Ltd. for permission to publish these results.

REFERENCE

WEBER, H., 1951, *Z. Phys.*, **130,** 392.

Some Characteristics of a Low Voltage Electron Immersion Objective

By W. W. H. CLARKE and L. JACOB
George Holt Physics Laboratories, University of Liverpool

Communicated by H. W. B. Skinner; MS. received 11*th February* 1952, *and in amended form* 25*th August* 1952

Abstract. An experimental investigation of the low-energy electron immersion objective in the range 20–70 v has revealed new features of interest in the study of beam distribution and characteristics. For example, the beam current shows a saturation effect over a considerable portion of the modulation range; further, beam-trimming electrodes absorb maximum and minimum fractions of the cathode current for particular values of the ratio α of anode to modulator voltage.

Using an arbitrary convention of measuring beam width, there appears to be a minimum beam radius at about 40 v for all beam angles, and an indication of a second minimum at higher energies and smaller beam angles. These and the marked change in distribution with voltage and with α are attributed to space charge repulsion effects.

The simple space charge theory breaks down for the lowest energies; the limit for production of an axial beam lying at about 20 volts. The trajectories in the presence of space charge indicate that even for the lowest emission densities, the 'crossover' may be considerably displaced.

---

§ 1. Introduction

The general behaviour of the electron immersion objective is now fairly well understood. This has followed a study of field patterns associated with the geometry of the system, together with an appreciation of their optical character as regards electron rays. The results of such investigations have proved important in the industrial applications of all kinds of electron devices which depend for their proper functioning on the production of an electron stream with defined characteristics. The function of the objective may thus be said to be to provide a beam of a certain diameter, electron density, energy and shape, the degree to which each of these parameters is specified depending on the application in view. This is achieved by the use of suitable, but variable, electrostatic fields produced between electrodes generally in the form of apertured discs, whose effect on the beam may control the extent of these factors to the desired amount.

Such 'guns' for medium and high voltages (up to 10 kv) have been described by Maloff and Epstein (1934), Law (1937), Langmuir (1937), Pierce (1940, 1944) and Jacob (1950). Their performance has been well summarized in review articles (Field 1946, Morton 1946). The principles associated with their design are usually based on laws derived empirically, though approximate theories have been proposed to explain their performance. These, however, suffer in most cases through neglect of the effects of space charge whose variable magnitude complicates the exact solution to an inordinate extent. Such a complicating feature when present at very low voltages (e.g. 20–70v) produces effects which completely alter the picture of beam formation, and introduce new patterns not encountered in the higher velocity regions.

The present work is directed to a study of some of the characteristics of an objective producing low-energy beams; the rough form of trajectories of specially selected electrons have also been calculated by a step-by-step method in which the space charge factor has been approximately assessed at each stage of the motion.

§2. THE EXPERIMENTAL SYSTEM

This consisted of an indirectly heated oxide cathode Ca, coated with a BaO/SrO mixture, as the electron source, a modulating electrode M and three electrodes A, B, C close to each other as shown in fig. 1. The latter can act either as a composite anode or may be used to produce variations in electron density angle, etc., in the beam emerging from the aperture in A. The emission was stabilized on the pump and the gun aged after 'gettering' and sealing off till the emission was again constant with time. A graphite coating on the walls of the tube consisting of three separate rings, each with its own connection, together with a series of concentric collecting discs at the end remote from C, all kept at the final accelerator voltage, served to collect the beam, while the return current to the cathode was separately recorded. The purpose of the system of collectors was to follow through changes in beam distribution when the space charge was altered.

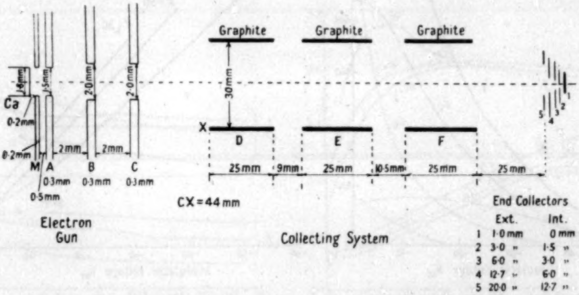


Fig. 1. Immersion objective and collector system.

To ensure that the stray field in the laboratory did not interfere with the beam formation in the objective, the axis of the gun system was carefully aligned along the resultant of the stray field and the earth's magnetic field. Delicate tests involving the use of concentric collectors were applied to detect and correct any asymmetry in the final beam. The adjustments were not considered satisfactory unless there was negligible deflection of the beam, and consequently negligible change in current to the collectors when the accelerator voltage was varied. Some idea of the effects of disturbing fields is given by the fact that a field of 0·03 oersted acting on a 20v beam over a length of 15 cm will deflect it through 1 mm.

The field in the cathode/modulator/anode region was determined by using enlarged models (× 10) of the electrode assembly in a set of separate experiments carried out in the electrolytic tank. This was subsequently used as the basis on which the trajectories were determined, both with and without the effect of space charge.

In order to assess gun performance under the simplest operating conditions the electrodes A, B, C were all kept at the final accelerator potential while the modulator potential was varied in value through a range from close to the cut-off point up to that of the cathode potential. The current taken by these electrodes, as well as that to the graphite rings and concentric collectors, was separately measured.

§3. THE GUN CHARACTERISTICS

The modulation characteristics of the usual immersion objective indicate that in all cases the total emission from the cathode passes through the aperture in the accelerator at all modulations, to be made use of in the focusing system following the gun. This no longer holds in the region of beam energies from 30–70 v. Here, while the cathode current I_k is substantial, a decreasing fraction of it reaches the focusing system as the modulator potential is carried towards the cathode potential. Further, there is a definite *falling away* of current emerging from aperture C (subsequently called beam current I_R) in the region between $V_M = -10$ and $V_M = 0$, and an actual reduction in axial flow. An example of this is shown in fig. 2 (*a*) for a 50 v beam; the currents taken by the A, B, C electrodes which constitute the accelerator are also plotted.

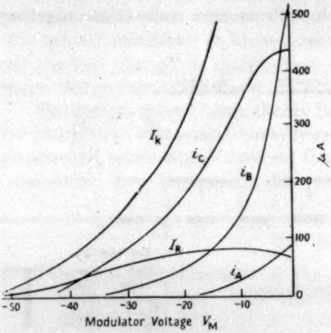


Fig. 2 (*a*). Beam characteristics.

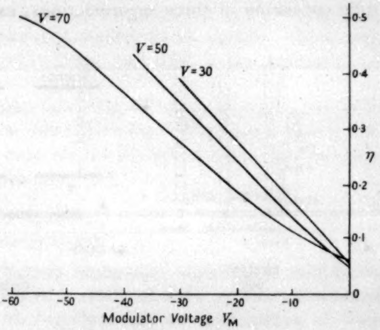


Fig. 2 (*b*). Efficiency of gun.

The normal effect of applying a negative bias to the modulator is to decrease the current in the beam, usually very rapidly over the first few volts of the modulation range and then more gradually, till a 'tail' develops towards the cut-off point. In the present case a negative bias of one-third the full modulation range, while suppressing the current to A, still allows considerable current to flow to B and C. This means that the beam shows a high divergence over the short region of 2 mm between A and B (and also between B and C) which does not disappear entirely till near the end of the modulation range. Though C appears to take most current over the major part of the range, there is a 'saturation' effect from about $V_M = -5$ to $V_M = 0$ which is most interesting. This effect, coupled with the nearly flat portion for the beam current I_B, indicates that the electron distribution in a 2 mm diameter core of the beam is more or less stationary at these high values of modulation close to the cathode potential. The distribution outside this core must be considerably altered in this range, since both the A and B electrodes, and particularly the latter, show substantial increases in current. It is obvious that space charge repulsion effects have completely altered the character of the beam structure at these low energies. This peculiar behaviour is reflected in the curves (fig. 2 (*b*)), showing the efficiency η of the gun (the ratio of beam to cathode currents). The fact that there are small variations in η at $V_M = 0$ is not surprising, because of the 'trimming' effects in the field-free space between electrodes B and C which were kept at anode potential. For any ratio of the modulation to anode voltage it is found that η increases with increasing accelerator voltage; further, for any given accelerator voltage η increases with increase in modulation bias towards the cut-off, due presumably to a continuous narrowing of the beam angle as occurs in an ordinary gun. This is what would be expected if the space charge were responsible for the change in electron beam structure across the beam.

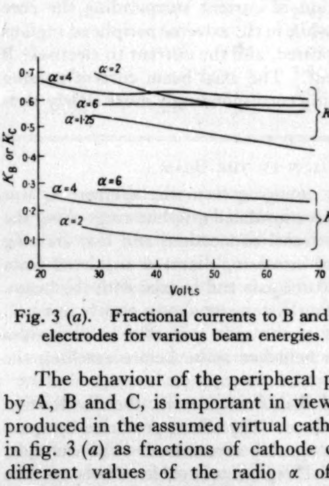


Fig. 3 (*a*). Fractional currents to B and C electrodes for various beam energies.

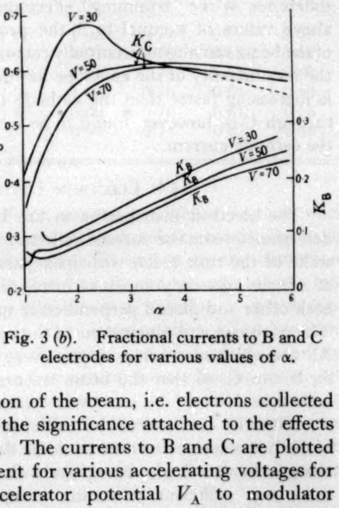


Fig. 3 (*b*). Fractional currents to B and C electrodes for various values of α.

The behaviour of the peripheral portion of the beam, i.e. electrons collected by A, B and C, is important in view of the significance attached to the effects produced in the assumed virtual cathode. The currents to B and C are plotted in fig. 3 (*a*) as fractions of cathode current for various accelerating voltages for different values of the radio α of accelerator potential V_A to modulator potential V_M, both with respect to the cathode. The values of V_M thus range from zero to the full cut-off voltage V_{BO}. The physical significance of α can be readily understood from the simple relationship $\theta_M = \theta_0\{(\alpha - C)/\alpha\}$ involving the angles θ_M at V_M and θ_0 at $V_M = 0$; C is a constant. In this case, as V_M increases the beam angle decreases, and it is well known that an approximately linear relation between these exists over about three-quarters of the modulation range beginning with $V_M = 0$. Hence over this range $\alpha(= V_A/V_M)$ varies from about unity to infinity, though most of the experimental results of the present work lie within the range $\alpha=1$ to $\alpha=6$. For any fixed V_A the beam angle is of course directly proportional to $V_{BO}-V_M$, and it would have been preferable to use this as a parameter in the measurements. This was, however, discarded owing to uncertainties in the value of V_{BO} arising from the long current tail with this special geometry. If the values of V_A and V_M be multiplied by the same constant factor k then it is simple to show that the trajectory, and hence the beam angle, is not affected unless the potential distribution is disturbed by the presence of the space charge. In this case the effects of potential depression may cause marked changes in distribution in the beam. An important purpose of the experiments was therefore to investigate how the electron distribution varied with electron velocity for various constant values of the beam angle, i.e. of the parameter α.

In fig. 3 (*a*) the tendency for the currents to B and C to decrease at the higher accelerator voltages is in accord with the view that the space charge repulsion effects are diminishing. Figure 3 (*b*), which examines the fractions of cathode current to B and C, expressed by K_B and K_C respectively in greater detail, reveals the existence of a maximum and minimum for particular values of α. Whereas K_C becomes almost constant for values of α greater than 3, K_B increases almost linearly with α, i.e. with increase in beam angle. This difference in the 'trimming' effect on the beam is interesting. It means that above values of α equal to 3 the proportion of current surrounding the core of the beam remains substantially constant, while in the extreme peripheral regions the rectilinearity of the envelope has disappeared, and the current to electrode B is increasing faster than the cathode current. The axial beam current passing through C is, however, found to be increasing correspondingly more slowly than the cathode current.

§4. ELECTRON DISTRIBUTION IN THE BEAM

The electron distribution in the beam emerging from the aperture C was determined from the current collected by three insulated graphite rings along the walls of the tube (each with its separate external connection) and that arriving at a series of nearly coplanar metal discs of increasing diameter insulated from each other and placed perpendicular to the tube axis and coaxial with the beam. The positions and dimensions of the collecting system are shown clearly in fig. 1. All the electrodes composing it were kept at the same potential as electrodes A, B and C, so that the beam traversed a field-free space before reaching the collectors.

With the currents to the collectors and their areas known, the distribution in the beam at any one time could be determined. The effects produced by space charge changes in the beam could thus be quickly ascertained. For example, there was sufficient resolution in the axial region to enable effects on the distribution to be followed through as the beam parameters were varied. Such experiments were carried out for different voltages on the accelerating system (A, B, C) and for various values of the ratio V_A/V_M. The table shows the values of the space charge factor $V^{3/2}/I_k$ for various values of α, i.e. for different beam angles at different accelerating voltages. It can be seen that this is reasonably constant in value, particularly for large beam angles. There are thus no anomalies in the space charge law for the total current, even down to the lowest energies used.

An interesting feature of the simple theory is that it assumes the electron trajectories do not cross, i.e. the current passing through the cross section embracing a particular trajectory does not vary with longitudinal axial distance. It is easily seen that where the potential depression is small the beam will remain homocentric if the current density is uniform. In the case of parallel trajectories in diverging beams it can readily be shown, on applying Gauss' theorem, that the force on the electron remains constant independent of radial height if the current density can be assumed to vary inversely with radius, a condition impossible to fulfil in practice.

The conditions in the beam forming portion of the gun are all important in following through the subsequent spread of the beam in its passage to the collectors. It is likely that, for large beam angles, we are dealing with a diverging core whose waist lies somewhere between the modulator and the A electrode. On the simple space charge theory it is possible to trace back the beam to an origin in the C aperture which, for the maximum current of 25 μA to the collectors (at 60v), corresponds to an area of radius 1 mm in the plane of C. As the beam angle is decreased this area contracts and the 'waist' of the beam will vary in both size and position, with a corresponding effect on the distribution.

Values of Space Charge Factor $V^{3/2}/I_k$ for Various Energies and Beam Angles

$V \backslash \alpha$	1	1·25	1·5	2·0	3·0	4·0	5·0	6·0
70	17·1	5·5	2·95	1·60	1·02	0·79	0·67	0·60
60	17·6	5·17	3·18	1·73	1·01	0·78	0·68	0·60
50	15·9	4·98	3·02	1·70	1·01	0·78	0·65	0·60
40	—	4·78	3·01	1·70	1·05	0·76	0·64	0·60
30	—	4·60	2·68	1·85	0·96	0·76	0·61	0·58
20	—	—	—	—	0·94	0·74	0·65	0·56

V is in volts, I is in μA.

A complicating factor which is difficult to assess without further work involves the aberrations in the lens system, since the distributions will obviously be affected by the chromatic and spherical error, as well as by the distortion. Further, it is known that their extent is governed by the strength of the lens system, so that an alteration of field may give rise to a variation in distribution arising from this cause. However, in the interpretation of subsequent distributions the lens aberrations will for the most part be neglected and the variations in beam distribution considered as arising from changes associated with the space charge, particularly in the region between the 'floodlit' aperture C and the concentric ring collectors at the end of the tube. Such an assumption is justified for high energy beams, which show no marked changes in the form of the distributions as the beam angle is varied by altering the focal properties of the electron gun (Jacob 1939).

A typical form obtained for the distribution curves at the collectors is shown in fig. 4 for varying accelerator voltages at $\alpha = 2$. The curves closely resemble gaussian distributions. The ordinates for any radial point roughly obey the 3/2 power law except for the lowest value of 30v, which shows a marked falling off of current density for all radial distances. This is shown more clearly in fig. 5 by plotting the current ratio for voltages of 70 and 30 to give a 'space charge line' (S.C.) for the whole system of electrodes taking current, from the moment the beam passes the anode A, for each value of α. The deviation for the near-axial region, corresponding with the inner collectors, is particularly noticeable, and the unexpected behaviour for electrode E represents a kind of periodic variation in density across the beam at the lowest energy.

Another feature of the curves is the very long tail associated with them even when the beam angle has been restricted by making α small, i.e. there is no sharp or clearly defined cut-off. This makes it impossible to assess the 'width' of the beam without the use of some convention which allows for the concentration of charge. It is proposed to adopt a purely arbitrary value of one-fifth the peak height, and to express all beam widths in terms of the width of the curve at this point. The results are shown in fig. 6, where the curves refer to different voltages and beam angles. Under these conditions there appears to be a minimum beam radius at about 40 v for all beam angles and an indication of a second minimum at higher voltages and small beam angles. Similar sets of curves with the same type of characteristic shape were also obtained by using conventions of $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ and $\frac{1}{8}$ of the peak height when assessing the widths. The very marked effect shown, corresponding to beam energies of about 30 v, could be repeated from day to day; this was done on many occasions in order to confirm its presence. An apparent anomaly of this kind might be questionable on two grounds: (i) possible

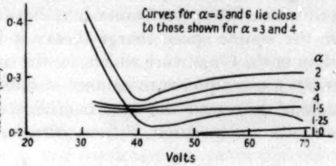


Fig. 6. Relation between beam width and beam energy.

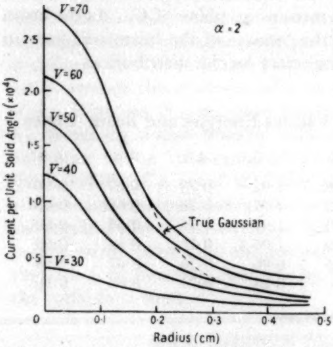


Fig. 4. Electron distribution to collectors for various beam energies.

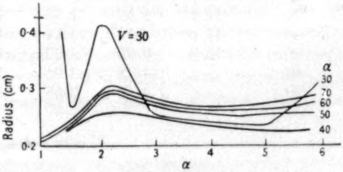


Fig. 7. Variation of beam width for various values of α.

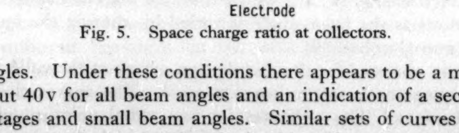


Fig. 5. Space charge ratio at collectors.

scattering effects due to gas present in the tube, (ii) secondary emission from apertures causing a variation of distribution across the beam. As regards (i), the processing on the pump, gettering, and clean-up action of the discharge when running over several months reduced the pressure below 10^{-8} mm. That such low pressures were attained was attested by the freedom from emission variations even for the smallest currents; further, at these pressures only about one out of 2×10^5 electrons leaving the cathode suffers a collision with a gas molecule; scattering effects must therefore be considered negligible.

As regards (ii), all electrodes were kept at anode potential. No variation in current to the end collectors could be detected when their potentials, individually and collectively, were varied by ±10% from the final anode potential. It was considered unlikely that any electrons except those striking the inner edge of the aperture in the forward direction should get into the beam. Of these, the number which get into the marginal regions of the beam will be limited by the trimming effects of electrodes A, B and C, and at these low energies the secondary emission ratio for thoroughly degassed electrodes is exceedingly small. Other factors tending to suppress such effects arise from (*a*) the reverse field as between the edge of the aperture and points on and near the axis in the vicinity of the centre of the aperture, (*b*) the space charge of the beam itself, particularly the negative field set up in the region between electrode C and the collectors; an estimate of its value shows it to be of the order of some tens of volts for the energies used. The distribution in the core of the beam as recorded must therefore represent the state of affairs existing in electron beams produced under the conditions stated.

The anomaly at about 30v is not surprising in view of later work which indicated the existence of oscillations in the plasma of the cathode region at about this energy. The simple electron optical feature of the focusing would lead to the expectation that, for constant α, the effect of the variation in V should not produce any change in the distribution, provided the space charge does not affect the paths. The marked change with voltage must hence be attributed mainly to the effects of space charge, which affects the distribution in the peculiar way shown by the curves. If the set is replotted in terms of α as in fig. 7, the same kind of irregularity is apparent. Here, only when $\alpha = 1$ does the beam width remain independent of energy, as it would do for a high energy beam at all values of α. For any particular energy the beam appears to have a maximum width for $\alpha = 2$, although the behaviour at 30v is irregular below this value of α. It then usually decreases or tends to become constant as the beam angle increases.

This behaviour must be the result of other important factors not previously considered, in combination with the space charge effect, e.g. the possible production of positive ions, both in the interelectrode space as well as in the 'field-free space', to the collectors. The variations of current density per unit collector current with voltage, for each individual collector, bear out the evidence that the simple space charge theory breaks down for low energy beams—the ratio for the central collector falls off at about three times the rate of that for the second concentric collector between 30 and 40v energy, and also shows a distinct minimum at about 65v energy. Below 20v the current to the central collector practically vanishes and the high values of beam radii at 30v indicate that this energy approaches the limit for the production of an axial beam with the gun under examination.

§5. The Field in the Objective

Field plots of enlarged models of the system comprising the objective were obtained in the electrolytic tank by two methods of measurement: (i) straightforward pick-up using a valve voltmeter as the measuring instrument, (ii) an improved method based on proportional measurement of the potentials, giving easier operation and greater accuracy in spite of fluctuations in the beat-frequency oscillator output voltage, i.e. potentials are measured in all cases as fractions of the supply voltage, which is placed across a potentiometer. An unbalance indicator, consisting of a long-tailed pair, rectifier and meter, was used in place of a galvanometer, since it has the great advantage of very high input impedance ($\sim 10^{10}$ ohms). The accuracy depends on the accuracy of calibration of the potentiometer and the sensitivity of the 'long-tailed pair' balance indicator. The anode loads were adjusted so that the gain of each triode was the same. Equal signals placed on the grids then gave minimum reading on the microammeter. Compensation for the slight difference in slope of the valve characteristics was achieved by the use of a condenser in the bridge circuit, as well as by running the valves at slightly different mean anode potentials to establish the correct balance condition at all inputs.

With an input of 20 v from the oscillator the sensitivity is such that a change in setting of the potentiometer by 0·001 of its value corresponds to an overall accuracy of within about 1% in the worst case. The axial potential distribution is shown in fig. 8(*a*). It conforms to an exponential law $V = Ae^{kz}$ over a considerable part of the cathode/modulator spacing. The variation of A and k with α is shown in fig. 8(*b*).

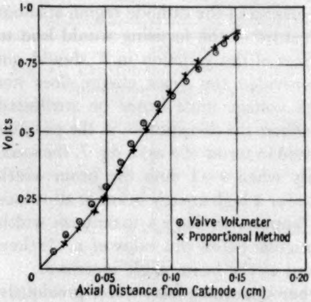


Fig. 8 (*a*). Axial potential distribution of objective.

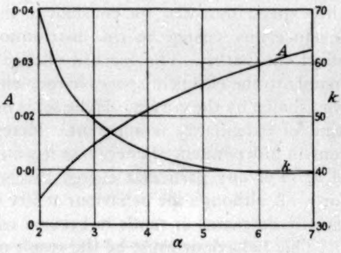


Fig. 8 (*b*). Variation of 'A' and 'K' with α.

§6. Trajectories in the Absence of Space Charge

The step-by-step method of Maloff and Epstein (1934) was used to trace trajectories, since it was considered unnecessary to apply the more recent refined and elaborate numerical techniques owing to uncertainties introduced by unknown complicating factors. Families of trajectories were computed for the various values of α, taking $V = 50$ v in each case. The initial velocity was assumed to be 0·1 v, or zero, for electrons directed parallel to the axis, while a value of 0·1 v was used for those leaving the centre of the cathode at 90° to the axis. The shapes of the paths in the focusing field are shown in fig. 9, in one case as far as its intersection with the axis, thus defining the plane of the 'crossover'. The size of the latter is seen not to vary greatly as the beam angle is increased above $\alpha = 2$, i.e. as the curvature of the field in the vicinity of the cathode is made weaker, which is somewhat surprising. Its location also appears to shift away from the cathode with increase in α. The position of the cathode image will not be sharply defined since it is bound to be greatly affected by the chromatic aberration at these low energies; this also is partly responsible for a crossover size of about 0·2 mm, an unusually large value with immersion systems.

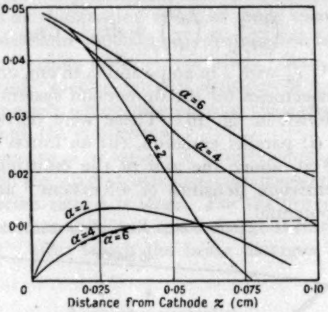


Fig. 9. Electron trajectories in the absence of space charge.

§7. Trajectories Involving Space Charge

An approximation to the type of trajectory taken by an electron when the space charge forces become appreciable in comparison with the focusing forces of the field may be obtained by using the step-by-step method and assuming (i) the space charge forces over cylindrical elements of the beam of length *dz* and radius *r* are given by simple application of Gauss' theorem, and (ii) these are additive vectorially to the electrostatic forces produced by the field in the gun. The effects of potential depression and all space charge effects which are non-radial are of course neglected here.

The outward force on an electron at radius *r* on the periphery of the envelope embracing a current *I* is thus derived from

$$md^2r/dt^2 = Ee = (2Q\mathbf{e}/r)\,dz, \qquad \ldots\ldots(1)$$

where *Q* is the charge in a cylindrical element of length *dz*. If *v* is the velocity, then

$$\frac{d^2r}{dt^2} = 2\frac{\mathbf{e}}{\mathbf{m}}\frac{I}{rv} \qquad \ldots\ldots(2)$$

at any *r*.

For the static field

$$\frac{d^2r}{dt^2} = -V'\frac{r}{\rho}\frac{\mathbf{e}}{\mathbf{m}}, \qquad \ldots\ldots(3)$$

where ρ is the curvature of the equipotential at the point considered. Thus

$$\frac{d^2r}{dt^2} = \frac{2I}{rv}\frac{\mathbf{e}}{\mathbf{m}} - \frac{V'r}{\rho}\frac{\mathbf{e}}{\mathbf{m}}. \qquad \ldots\ldots(4)$$

Now $v=(z_1-z_0)/t_{01}$ for the step 0 to 1, where t_{01} is the time taken to traverse the path between 0 and 1, so that

$$\frac{d^2r}{dt^2}=\frac{2It_{01}}{r(z_1-z_0)}\frac{\mathbf{e}}{\mathbf{m}}-\frac{V'r}{\rho}\frac{\mathbf{e}}{\mathbf{m}}, \qquad \ldots\ldots(5)$$

The velocity at the end of step 1 is given in terms of the velocity at the beginning by $v_1=v_0+(\text{acceleration}\times\text{time})$; then

$$\left(\frac{dr}{dt}\right)_1=\left(\frac{dr}{dt}\right)_0+2\frac{\mathbf{e}}{\mathbf{m}}\frac{I}{r_0}\frac{t_{01}^2}{z_1-z_0}-\frac{r_0V''}{2}t_{01}\frac{\mathbf{e}}{\mathbf{m}}. \qquad \ldots\ldots(6)$$

This can be transformed for use with practical units and a step of 0·01 cm to give

$$\left(\frac{dr}{dt}\right)_1=\left(\frac{dr}{dt}\right)_0+At_{01}^2-Ct_{01}, \qquad \ldots\ldots(7)$$

where $A=3{\cdot}15\times10^{29}\,I/r_0$, with I in amps and r_0 in cm, and $C=8{\cdot}765\times10^{14}\,r_0V''$.

A rough set of trajectories for the immersion system at 50 v in the presence of space charge is shown in fig. 10. These were derived from the preceding analysis and assume (i) parallel emission, (ii) an initial energy of 0·1 v, (iii) an initial height of 0·05 cm above the axis of the emitting cathode. They were traced for average emission densities of 4·5 mA cm⁻² and with the modulator electrode at cathode potential.

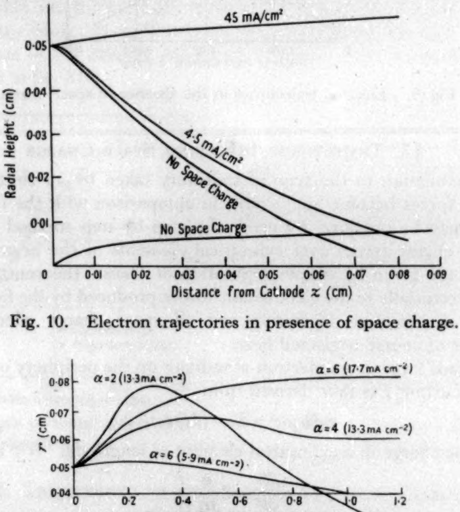


Fig. 10. Electron trajectories in presence of space charge.

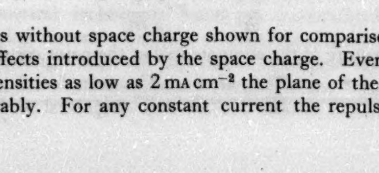


Fig. 11. Trajectory plots for various values of α.

The trajectories without space charge shown for comparison give some idea of the profound effects introduced by the space charge. Even for average total cathode current densities as low as 2 mA cm⁻² the plane of the crossover may be displaced considerably. For any constant current the repulsion effect is more pronounced the closer the emitting point is to the centre of the cathode, as is to be expected, and if close enough can even result in a defocusing effect which may prevent part of the beam from passing through the aperture in the modulating electrode. Under certain conditions of current density the balance of forces may be such as to produce a parallel beam over small regions of the field; this soon gives way to divergence, weak or strong, depending on the extent of the modified curvature introduced by the space charge.

The plots for various values of α with the same system are interesting. These are shown in fig. 11 for an originating height of 0·05 cm. Here again there is a displacement of the crossover plane with current density; there is a further displacement with α. The dominating effect of space charge repulsion in the region between the cathode and modulator is appreciable for small values of α, and there is a definite converging effect on the whole. As α increases, this effect does not appear so striking; average current densities of only some 10 mA cm^{-2} give place to a diverging effect which may prevent the peripheral portion of the beam from passing through the gun system.

Generally, the electrons tend to remain at a fairly constant radius during the early stages of the motion for current densities corresponding with experimental values when the modulation ratio α is large. For the lower values of α there is the constant interplay between field and space charge forces which results in the formation of the 'waist', after which the beam diverges in its passage to the collecting electrodes.

Acknowledgment

The authors wish to express their thanks to Professor H. W. B. Skinner for his warm interest in this work.

References

Field, L. M., 1946, *Rev. Mod. Phys.*, **18**, 353.

Jacob, L., 1939, *Phil. Mag.*, **28**, 81; 1950, *Proc. Phys. Soc.* B, **63**, 75.

Langmuir, D. B., 1937, *Proc. Inst. Radio Engrs., N.Y.*, **25**, 977.

Law, R. R., 1937, *Proc. Inst. Radio Engrs., N.Y.*, **25**, 954.

Maloff, I. G., and Epstein, D. W., 1934, *Proc. Inst. Radio Engrs., N.Y.*, **22**, 1386.

Morton, G. A., 1946, *Rev. Mod. Phys.*, **18**, 362.

Pierce, J. R., 1940, *J. Appl. Phys.*, **11**, 548; 1944, *Ibid.*, **15**, 721.

A Miniature Helium Liquefier-Cryostat of Cascade Type

By P. F. CHESTER and G. O. JONES

Department of Physics, Queen Mary College, University of London

MS. received 15*th December* 1952

Abstract. A miniature cascade helium liquefier-cryostat is described which will cool from room temperature to 4·2°K in half an hour and enables continuous experiments to be carried out at or near this temperature, or at intermediate temperatures, for as long as required. It consumes only liquid oxygen and hydrogen from high-pressure cylinders, is extremely economical in its consumption of these materials, and incorporates a number of new features which make its operation almost automatic after the starting-up period.

§1. INTRODUCTION

THE difficulties of carrying out complicated experiments at liquid helium temperatures are aggravated if the experimenter is dependent on others for the supply of liquid helium unless it can be kept available at all times. Such a service would be costly and has not yet been provided in any laboratory in this country. A complete solution is provided in principle if each experimenter can make his own liquid helium using only materials always available. This approach is particularly promising if the liquefier and the experimental working space, or cryostat, can be combined in the same envelope, because transfer losses are then avoided and only a very small quantity of helium need usually be liquefied. This miniature technique, developed by Ruhemann (1930) and Simon (1936), has been widely employed. In spite of its obvious potential advantages, the cascade principle, in which all stages of liquefaction are combined in one envelope, has seldom been applied with the miniature technique. The present paper describes a miniature cascade liquefier which differs from previous designs in that particular attention has been directed to reducing the starting-time and to simplicity and continuity of operation, in order that the need for each experimenter to run his own liquefier should not itself add to the difficulties of the experiment. Special attention has been paid also to the question of flexibility so that experiments of many different kinds can be performed with the minimum of adjustment.

§2. PRINCIPLES OF THE DESIGN

The miniature cascade liquefiers described by Rollin (1936) and Schallamach (1943) employ Joule–Kelvin expansion at the liquid hydrogen stage and the Simon expansion method at the liquid helium stage. We have employed the Joule–Kelvin method at both stages because its use in the helium liquefaction reduces the starting-time, permits the liquefaction of further quantities of helium at any time without warming to hydrogen temperatures, and makes for easy adaptability to a wide range of experimental applications. The greater difficulties in the Joule–Kelvin method arising from contamination have been overcome by the use of a small oil-free gas-tight compressor of new type (Jones 1951) working in a closed helium circuit which includes a dry gas-holder in the form of a large rubber bellows. (We are indebted to Professor J. F. Allen for kindly supplying us with information about such gas-holders.)

The design and circuits of the liquefier are shown schematically in fig. 1. There are four stages of cooling. The main vacuum chamber A is covered to a depth of several centimetres by liquid oxygen in an open Dewar vessel. The inner container B is maintained at 62°K by liquid oxygen boiling in it at a pressure of about one cm. High-pressure hydrogen, pre-cooled to 62°K, expands into vessel C, in which a pressure of about 20 cm abs. is maintained by a rotary pump at the outlet of the return exchanger. The liquid hydrogen in vessel C serves to cool the high-pressure helium to about 16·5°K before Joule–Kelvin expansion into vessel D. Vessel C also carries a radiation shield E for the protection of the helium stage. Constructional details will be found in the Appendix.

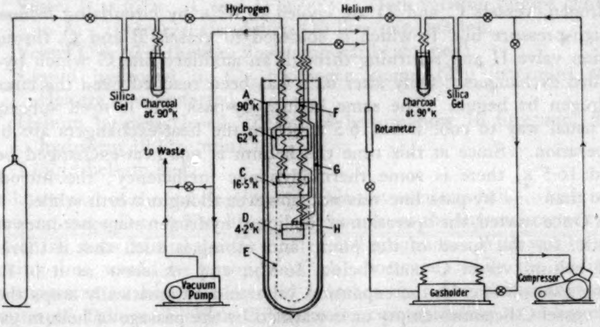


Fig. 1.

For reasons of economy it was decided to use liquid oxygen rather than liquid air for pre-cooling, in spite of the fact that lower temperatures can be reached with liquid air, and to make as full use as possible of the cylinders of hydrogen by continuing to use them down to, say, 35–40 atm. To obtain a reasonable yield of liquid hydrogen at the lower cylinder pressures a somewhat low pre-cooling temperature is required, and this is achieved, using liquid oxygen, by the introduction of the second pre-cooling stage B. Since this stage is well insulated and is used only to cool gas from 90° to 62°K, the required degree of vacuum can be maintained by quite a small pump. (We have found Aroclors 1248 and tritolyl phosphate satisfactory as non-inflammable pump oils for use in pumping oxygen.)

With the small flows of hydrogen necessary, it is possible to maintain the liquid hydrogen at a pressure of 20 cm abs. or less, again using quite a small pump. The rate of flow of helium is adjusted so as to evaporate the liquid hydrogen at about the same rate as it is formed and, because *all* the returning hydrogen is heat-exchanged with the incoming hydrogen between 16·5°K and room temperature, maximum refrigeration at 16·5°K for a given flow of hydrogen is achieved. It is interesting to compare the present design with that of its closest relative, the much larger Joule–Kelvin cascade liquefier at the Royal Society Mond Laboratory, Cambridge (Ashmead 1950), where one liquid air stage and two liquid hydrogen stages are employed as compared with our two (liquid oxygen) and one stages respectively. The reasons for our introduction of a second liquid oxygen stage have been explained. The use in the large Cambridge liquefier of a second liquid hydrogen stage was necessary to avoid the use of an impracticably large pump.

The following special features are incorporated in the present design:

(i) An automatic valve dispenses liquid oxygen (via a gauze filter) from above vessel A to vessel B as required to keep the level in B constant. This valve is a simplified version of a type already described (Jones 1948) and is actuated by a single vapour pressure operated bellows connected to a bulb situated inside B at the desired level. In practice, stage B thus requires no attention from the operator.

(ii) In order to achieve a short starting-time without wasteful consumption of hydrogen, vessels C and D are cooled to 62°K by circulating ***helium*** in its own high-pressure line F, which is soldered to vessels B and C, through the expansion valve H and returning through an auxiliary line G which by-passes the return exchangers. Only after 62°K has been reached need the circulation of hydrogen be begun. The same helium by-pass line is used subsequently in the usual way to cool D to 16·5°K before the heat-exchangers are brought into operation. Since at this time the helium is not heat-exchanged between 90° and 16·5°K, there is some thermodynamic inefficiency; the introduction of more than one by-pass line was not however thought worth while.

(iii) Once started, the operation of the liquid hydrogen stage becomes entirely automatic, for the speed of the pump and tubing is such that if there is no flow of helium vessel C cools below 16·5°K, and as soon as it is full the incoming gas solidifies at the expansion valve and automatically stops the flow. Only if vessel C becomes empty or is warmed by the passage of helium gas does the 'freezing valve' open and allow more hydrogen to pass. In practice it is found that this stage again can be left quite unattended for, say, a whole day. It will be noted that this innovation automatically restricts the production of liquid hydrogen to that quantity necessary for the liquefaction and protection of the helium required for the experiment. In conjunction with the feature mentioned in the preceding paragraph, it limits the consumption of hydrogen gas to the minimum quantity actually necessary.

Once the desired quantity of helium has been liquefied the liquefier as a whole requires no attention except the addition (about twice an hour) of liquid oxygen to the (open) Dewar vessel.

A number of other points may be noted briefly. With silica gel and charcoal traps in the hydrogen circuit it has been found possible to use a fixed (pre-set) expansion valve (after Ruhemann 1930), with consequent saving in trouble and complication in the design. The flow is adjusted by varying the pressure before the expansion valve, since a satisfactory liquefaction yield can be obtained over a wide range of pressures. The situation is somewhat different in the helium circuit because there is a clear optimum pressure for the liquefaction of helium between 20 and 25 atmospheres at our starting-temperature, and the variations in pressure required to vary the flow between the desired limits during a run may extend well outside this region. It is therefore necessary to have the helium valve under positive control, and a screw and screwdriver valve similar to that described by Schallamach (1943) is employed. Pressure gauges and vapour pressure thermometers are fitted at appropriate points in the liquefier, and a Rotameter in the low-pressure line indicates the rate of flow of helium. At present silica gel and charcoal traps are included in the helium circuit as a precautionary measure.

§ 3. PERFORMANCE

The following figures refer to a timed run conducted in order to obtain performance data.

Time (min)	
0	Whole apparatus at room temperature. Filling with liquid oxygen started. Oxygen pump (250 litre/min) started, and helium circulation (40 litre/min) started by switching on compressor.
6	Vessel B at 62°K. C and D at 90°K. High vacuum pumps connected to A. Hydrogen circulation (30 litre/min) started.
9	Vessel C at 62°K (helium flow temporarily stopped).
14	Liquid hydrogen in C. Hydrogen pump (250 litre/min) started. Helium flow restarted.
22	Helium by-pass closed. (Flow of helium now 16 litre/min, flow of hydrogen 10 litre/min)
32	Liquid helium in D.

The latter part of the starting-time (in this test over one half), which is taken up in cooling stage D to 4·2°K, depends of course on the thermal capacity at D. In this test the mass (of brass) at D totalled 160 grammes. With smaller masses the starting-time would be correspondingly shortened, while the addition of further apparatus to be cooled would lengthen it. The subsequent rate of liquefaction of helium is about 2·5 cm^3/min, corresponding to a liquefaction coefficient of 0·11—about 85% of the theoretical yield. The rate of evaporation of liquid helium will normally vary between 2 and 5 cm^3/hour, according to the type of experiment being carried out, so that the quantity of liquid produced in ten minutes' additional running will usually suffice for 5 to 10 hours' working. When more is required it is necessary only to re-start the flow of helium gas for a few minutes.

Approximately 400 litres of hydrogen (measured as free gas) are used in cooling the liquefier to 4·2°K, 35 litres for every 10 cm^3 of liquid helium produced, and a further 40 to 50 litres per hour subsequently in making good evaporation losses. About 4·5 litres of liquid oxygen are used in pre-cooling the liquefier and producing 50 to 100 cm^3 of liquid helium, and about 0·4 litre per hour subsequently in making good evaporation losses. It will be noted that in the above timed run the flow of hydrogen was started earlier than necessary, thus reducing the starting-time but somewhat increasing the consumption of hydrogen. Under these conditions two full days' work, each of up to 10 hours' duration and requiring 50 to 100 cm^3 of liquid helium, may be carried out using one cylinder of hydrogen of standard size (165 ft^3). If, in starting up, vessels C and D are cooled to 62°K by the circulation of helium only, as described, the starting-time is increased by about 10 minutes, but now three days' work may be expected from one cylinder of hydrogen. The consumption of liquid oxygen is about 8 litres in a full day. The quantity of helium in the circuit need not exceed 200 litres.

ACKNOWLEDGMENTS

We acknowledge gratefully the help and facilities given us by Professor H. R. Robinson (Head of the Department of Physics), the careful work of Mr. W. A. G. Baldock of the departmental workshop and the assistance of Mr. H. W. Hunter.

We are grateful to the Department of Scientific and Industrial Research for a maintenance allowance awarded to one of us and for a special grant in support of the work, to the Central Research Fund of the University of London for the loan of apparatus, and to Imperial Chemical Industries Ltd. for a grant-in-aid.

APPENDIX

By P. F. Chester, J. A. Hulbert, G. O. Jones and D. L. Martin

CONSTRUCTIONAL DETAILS

Three helium liquefiers of the type described are now in existence. The performance data already given were obtained with the first model to be built, but a number of constructional improvements have been incorporated in the later models, which are described below and illustrated in fig. 2. The letters used in this figure are consistent with those already used in the main part of the text.

The dimensions were chosen to allow the liquefiers to fit inside large glass Dewar vessels (50 cm length, 10 cm diameter) of a type now commercially available. For greater reliability and rapidity of cooling it was decided to use all-metal construction. The outer vacuum case A and the inner vessels B, C, D are of brass and the wide vertical tubes of stainless steel. All joints or seals which are required to be gas-tight are hard-soldered, except that the expansion valves are soft-soldered into position, and the outer vacuum-case is joined to the top-plate by an easily demountable seal of Wood's metal. The liquid oxygen control valve is soft-soldered into position from outside the vacuum chamber. Short tubes of stainless steel are made integral with the top-plate, so that small tubes passing into the vacuum space may easily be hard-soldered into position without heating the plate. An eccentric arrangement of vessels and tubes has been adopted. This allows a tube of 1 in. diameter and several smaller tubes (including a number of 'spare' tubes added to increase the flexibility of the apparatus) to be passed straight down from outside the liquefier to its coldest parts. All such vertical tubes are reduced in wall-thickness over parts of their length.

For simplicity in construction the heat-exchangers are made of pairs of cupro-nickel tubes (H.P. tubes: $\frac{3}{32}$ in. o.d., 26 s.w.g.; L.P. tubes: $\frac{1}{8}$ in. o.d., 26 s.w.g.; length 1–2 m) hard-soldered together side by side. Cylindrical sleeves serve to connect the separated ends of the exchangers to other parts. Cooling coils for the high-pressure gases are soft-soldered to the outside of vessels B and C. The by-pass return path for the low-pressure helium is provided by the $\frac{1}{4}$ in. tube which sheaths the control rod of the expansion valve; this tube can be used as a low-resistance pumping tube to lower the temperature of the helium below 4·2°K. The vessels containing liquid hydrogen and liquid helium are fitted with outlets (normally sealed) in their bases to which connections to other vessels may be made if necessary. If such a vessel is attached to the 1 in. tube mentioned above, liquid can be decanted into it, and specimens up to nearly 1 in. diameter cooled to 20·4° or 4·2°K by inserting them vertically from outside the apparatus. A fraction of the helium gas circulated may be drawn off either before or after the expansion valve for the purpose of cooling a specimen in the vacuum space. All stages carry drilled copper stubs for the purpose of thermally 'fixing' suitable points along the length of small tubes or wires which may be passed down from room temperature for use in a particular experiment. The liquid hydrogen vessel is attached to a horizontal plate carrying a threaded flange upon which is screwed the radiation shield, thermal contact being made by a film of vacuum grease. This is not intended to form a vacuum-tight joint; as may be seen from the section YY, fig. 2, there are several openings in this plate which allow the passage of tubes or wires into the space inside the shield. We have deliberately avoided a multiplicity of vacuum joints by suspending all the inner vessels in a single vacuum space.

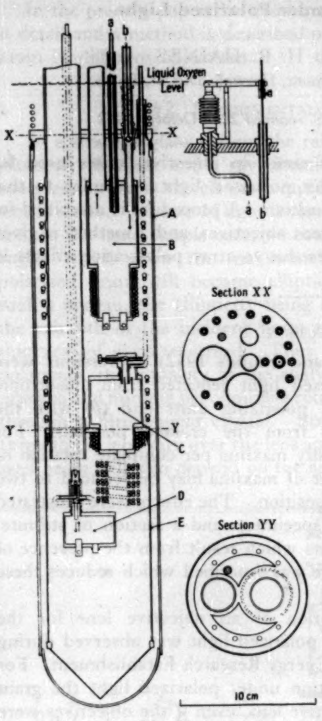


Fig. 2.

In the figure a simple vessel is shown at D; alternative apparatus may be substituted by breaking and re-making a single soft-soldered seal at the expansion valve, thus permitting the use of a variety of experimental arrangements. We have used, for instance, the device of Daunt and Mendelssohn (1948) which renders the expansion valve remote from the liquid helium so as to reduce evaporation, and a charcoal adsorption cryostat suitable for work between 4° and 10°K.

REFERENCES

ASHMEAD, J., 1950, *Proc. Phys. Soc.* B, **63**, 504.
DAUNT, J. G., and MENDELSSOHN, K., 1948, *J. Sci. Instrum.*, **25**, 318.
JONES, G. O., 1948, *J. Sci. Instrum.*, **25**, 239; 1951, *Ibid.*, **28**, 313.
ROLLIN, B. V., 1936, *Proc. Phys. Soc.*, **48**, 18.
RUHEMANN, M., 1930, *Z. Phys.*, **65**, 67.
SCHALLAMACH, A., 1943, *J. Sci. Instrum.*, **20**, 195.
SIMON, F., 1936, *Proc. VIIth Int. Congr. of Refrigeration*, p. 253.

The Effect of Strain in Objective Lenses used for Microscopical Examination of Metals under Polarized Light

By B. W. MOTT and H. R. HAINES
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

Communicated by L. F. Bates; MS. received 22nd October 1952

Abstract. It is shown that the effect of strain in objective lenses used for microscopical examination of metals under polarized light is similar to that of rotation of the analyser from the cross position. A procedure is described for assessing the relative strain effects in different objectives, and a method is given for compensating for the various anomalies due to stray polarization effects in the microscope.

§ 1. Introduction

In an earlier paper (Woodrow, Mott and Haines 1952) expressions were derived for the intensity of the analysed light reflected from anisotropic absorbing crystals, (*a*) under crossed polarizing units and (*b*) with the analyser rotated through a small angle from the crossed position. For condition (*a*) it was shown that four intensity maxima per complete rotation of the specimen are obtained, but the number of maxima may be reduced to two as the analyser is rotated from the crossed position. The effects were illustrated by experimental curves on various metal specimens and a section of stibnite. Mention was made of the undesirable effects which result from the presence of strain in the objective lens, and a procedure was described which reduces these effects to a minimum.

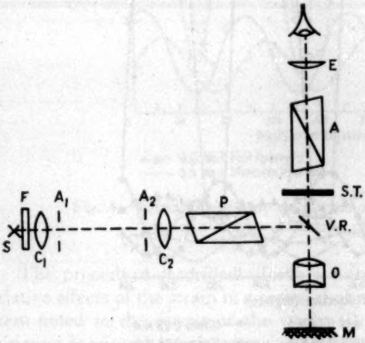
The significance of the strain condition of an objective lens for the examination of metal surfaces by reflected polarized light was observed during the initial work on uranium at the Atomic Energy Research Establishment. For a specimen suitably polished for examination under polarized light the grain contrast varied considerably with the objective lens, even if the objectives were of the same numerical aperture. (When used under ordinary illumination all the lenses were satisfactory.) The grain contrast also showed marked variations on rotation of a given objective, and this is illustrated in figs. 1 (*a*), (*b*) and (*c*) (Plate), which were taken on an electrolytically polished uranium specimen for three settings of an 8 mm objective, the photographic procedure being the same in each case.

Berek (1937), Von Schwarz (1931) and Capdecomme and Orcel (1941) have drawn attention to the unsuitability of strained objectives for polarized light work but have not elaborated on the effects of the strain. Neuerberg (1948) has described the effects of strain on the isogyr figure seen in the back lens of an objective with the eyepiece removed. The isogyr marks the vibration directions of the polarizer and analyser units and, in a perfect system, should take the form of a dark cross. Due to stray polarization effects, however, the figure is rarely a cross, and usually consists of two hyperbolae which change their form as a strained lens is rotated in its mount between crossed polarizers. Microscope manufacturers have now recognized the anomalies that can arise from strained objectives and will supply special objectives for polarized light work which are relatively strain-free.

In the present paper the consequences of strained objectives are considered in detail and a method is described of comparing the relative magnitudes of the strain in different objectives.

§ 2. EXPERIMENTAL PROCEDURE AND RESULTS

In a strained glass system the refractive index depends on the stress in the plane of vibration, and a ray vibrating in a plane parallel to the direction of a compression stress travels more rapidly than a ray vibrating in the direction of a tensile stress. A beam of light is resolvable into two plane polarized components and, in the most simple case, we can assume two 'principal' directions at right angles which show the maximum difference of refractive index. A plane polarized beam will become elliptically polarized after transmission through such a system, the ellipticity being dependent on the magnitude of the stress, the azimuth of the incident beam relative to the principal directions and the thickness of glass traversed. Thus, if the objective is rotated between crossed polarizers, the intensity of the analysed light would be expected to show four maxima and minima per complete rotation in the same manner as a random section through a birefringent crystal. Since the stress system and the optical path through the lens vary over the lens aperture, the net interference to a plane polarized beam will also depend on the aperture of the field iris of the microscope.



S = Source
F = Filter
C_1 = Condenser
A_1 = Lamp iris
A_2 = Field iris
C_2 = Condenser
P = Polarizer
V.R. = Vertical reflector
O = Objective (refracting or reflecting)
M = Metal specimen
S.T. = Sensitive tint
A = Analyser
E = Eyepiece or photomultiplier tube

Fig. 2. Schematic diagram of apparatus.

The apparatus used was the same as that described previously (Woodrow, Mott and Haines 1952) and is shown schematically in fig. 2. The rotating stage and objectives were centred with respect to the optical axis of the microscope; the analyser, eyepiece and objective were removed and the polarizer was rotated to give maximum intensity in the microphotometer unit inserted in the eyepiece tube when a highly polished, stainless steel specimen was placed on the stage.

In this position the plane of polarization was in the plane of incidence of the zinc sulphide coated glass slip used for the vertical reflector. The analyser was then inserted and crossed with the polarizer by rotating to give minimum intensity with the lamp iris A_1 and field iris A_2 closed to minimum aperture. A special objective mount was constructed such that the lens could be freely rotated about the optical axis without affecting the centring. If a strain-free objective is now inserted, due to loss of light by absorption and reflection, the photometer reading should be slightly reduced. In general it was found to increase and to vary markedly on rotation of the objective, going through four maxima and minima per complete rotation, the maxima occurring at approximately 90° intervals.

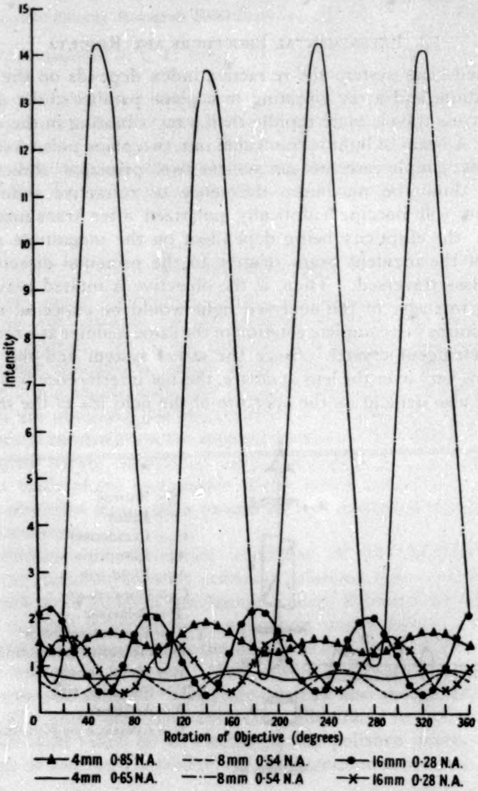


Fig. 3. Curves of intensity against lens rotation for various objectives.

Typical curves obtained on rotation of six transmission objectives are shown in fig. 3. The lowest intensity change was given by a 4 mm focal length lens, specially supplied strain-free by the manufacturers, for which the photometer readings varied from 0·74 to 0·85. An 8 mm focal length objective which had been particularly troublesome for visual work gave a variation from 0·75 to 14·2. The variation for a Burch-type reflecting objective was 0·72 to 0·80, which is comparable with the special strain-free transmission lens. The range of intensity variation was found to increase with the setting of the field aperture A_2, and the results for two objectives are given in fig. 4.

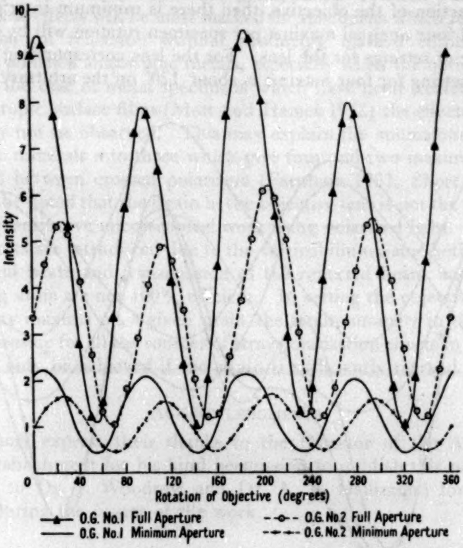


Fig. 4. Effect of aperture on curves of intensity against lens rotation for two 16 mm refracting objectives (0·28 N.A.).

The procedure described affords a suitable qualitative test for assessing the relative effects of the strain in a series of objectives. Two characteristics of lenses were noted in the course of the work: (i) As would be expected, the intensity variation increased as a lens was screwed tighter in the mount, and to keep the strain effects to a minimum it was necessary to control the degree of tightening. (ii) There is a recovery effect for an objective with time as shown by the fact that the difference between the maxima and minima intensities on lens rotation decreases over a period of several hours after mounting and, in addition, the position of the maxima and minima change. This can be attributed to the complex nature of an objective since, if part of the lens system recovers from strain to a greater degree than the rest, a shift may occur in the effective principal diameter.

The strain in the objective has an important effect on the shape of the intensity–azimuth curve for an anisotropic specimen. This was shown by taking an electrolytically polished bismuth section and determining the intensity–azimuth curves for different settings of the objective in its mount. The results are shown in fig. 5, in which the numerals by the side of each curve represent an approximate angle of rotation of the lens from an arbitrary position with an accuracy to about ± 5°. The effect of the strained objective is to reduce the number of intensity maxima from four to two and to change the grain contrast as for rotation of the analyser from the crossed position (Woodrow, Mott and Haines 1952). If the plane of vibration of the polarizer approximates to a principal direction of the objective, then there is minimum interference by the objective and four unequal maxima per specimen rotation will be obtained over a small range of settings for the lens. For the lens corresponding to the curves in fig. 5 the setting for four maxima is about 120° on the arbitrary scale chosen.

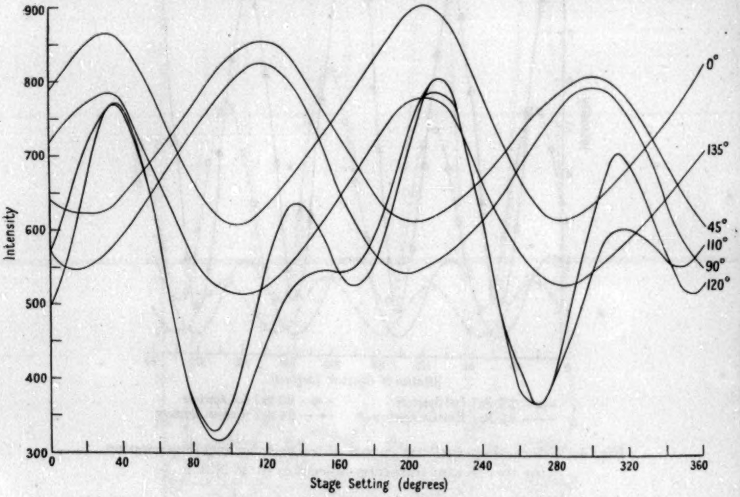


Fig. 5. Effect of rotation of the objective on intensity curve for bismuth.

When the principal direction of the lens coincides with the plane of vibration of the polarizer there will be no interference to the incident beam due to strain in the lens, but there will be a slight effect on the reflected beam if rotation of the plane of polarization occurs on reflection at the specimen. Thus the optimum setting of the lens is slightly offset from the above position, and such that its effect on the incident beam is corrected by its effect on the reflected beam. The position for correction will therefore depend on the angle of rotation and the ellipticity resulting at the specimen, and will vary both with the orientation of the grain and the anisotropic properties of the material.

§ 3. REMARKS

It has been shown that the effect of strain in objective lenses used for the microscopical examination of metals under polarized light is similar to that of rotation of the analyser from the crossed position. The number of intensity maxima for a complete rotation of the specimen may be reduced from four to two, and the grain contrast observed in a multigrained structure is markedly affected. It is possible, however, to set the objective in a position where four equal intensity maxima are obtained for a given grain, but this setting will vary with the grain orientation. Rotation of the lens from this position at first increases the grain contrast but, if the strain is sufficiently great, further rotation will decrease the contrast. These effects will be most marked for specimens which rotate the plane of polarization on reflection without producing marked ellipticity, i.e. for minerals and polished anisotropic metals, especially when the optical anisotropy is weak. In the case of metal specimens which have been treated to give etch pits or anisotropic surface films (Mott and Haines 1952) the effects of a strained objective may not be observed. This may explain the anomalous classification of anisotropic minerals into those which give four and two maxima respectively when rotated between crossed polarizers (Farnham 1931, Short 1940).

It should be noted that the strain in the objective lens is not the only source of anomaly in quantitative microscopical work using polarized light. Rotation and ellipticity effects are introduced due to the vertical illuminator both on reflection of the incident beam and transmission of the reflected beam, and, in addition, the polarizing units are not 100% efficient. In setting the objective to give four equal intensity maxima on a given grain the strain anomaly in the objective is used to compensate for all the sources of stray polarization effects in the apparatus, and this can only be achieved if the strain is sufficiently marked.

ACKNOWLEDGMENTS

The authors express their thanks to the Director of the Atomic Energy Research Establishment for his kind permission to publish this paper. We are also grateful to Dr. J. Woodrow and Dr. A. F. Hallimond for many useful discussions during the course of the work.

REFERENCES

BEREK, N., 1937, *Fort. Min. Krist. Petr.*, **22,** 1.
CAPDECOMME, L., and ORCEL, J., 1941, *L'Optique*, **20,** 47.
FARNHAM, C. M., 1931, *Determination of Opaque Minerals* (New York: McGraw-Hill).
MOTT, B. W., and HAINES, H. R., 1951, *Research, Lond.*, **4,** 24, 63.
NEUERBERG, G. J., 1948, *Amer. Mineralogist*, **33,** 496.
SHORT, N. M., 1940, *Geological Survey Bulletin* No. 914 (Washington: U.S. Department of the Interior).
VON SCHWARZ, M., 1931, *Metallurgia, Manchr.*, **4,** 180.
WOODROW, J., MOTT, B. W., and HAINES, H. R., 1952, *Proc. Phys. Soc.* B, **65,** 603.

The Reflection of Radio Waves from an Ionized Layer having both Vertical and Horizontal Ionization Gradients

By R. P. WALDO LEWIS
Physics Department, University College of Swansea

Communicated by W. J. G. Beynon; MS. received 2nd July 1952

Abstract. The approximation of ray optics is used to calculate the effects of simultaneous vertical and horizontal ionization gradients on the paths of radio waves in the ionosphere. Cases of practical importance are considered.

---

§ 1. Introduction

In general, the methods of calculating radio wave transmission paths in the ionosphere assume that the ionization density varies only in the vertical direction. Methods for calculating such paths have been given by Appleton and Beynon (1940, 1947), Smith (1937).

However, it is well known that large and persistent horizontal gradients sometimes exist, and under such conditions appreciable distortion of the transmission paths may be expected as has been shown by Rawer (1951). In this paper we consider short wave propagation in two simple cases involving both vertical and horizontal ionization gradients. We shall employ the method of ray optics, neglecting the effect of the earth's magnetic field and of ionospheric absorption, and consider only the case in which the plane of incidence contains the direction of the horizontal gradient. Typical results will be illustrated in the diagrams.

§ 2. Horizontal and Vertical Ionization Gradients

Type (*a*)

Type (*a*) is defined as a case in which the thickness of the ionized layer and the height of its lower edge are constant in the horizontal direction, and in which the electron density distribution is parabolic in every vertical section, and varies linearly in the horizontal direction.

Setting the origin of coordinates at the point where the ray enters the layer, we take the y axis vertical and the positive x axis along the direction in which the electron density (at a given y) increases. We denote the semi-thickness of the layer at $x=0$ by y_{m0}, and it will be convenient to write $Y=y/y_{m0}$, $X=x/y_{m0}$. The type of ionization distribution considered will then be represented by the equation $N=N_0(2Y-Y^2)(1+\lambda X)$ where N_0 is the maximum electron density in the vertical plane at $x=0$, and λ is a constant. The refractive index μ for frequency f is then given by

$$\mu^2=1-k^2(2Y-Y^2)(1+\lambda X) \qquad \ldots\ldots(1)$$

where $k^2=e^2N_0/\pi mf^2$ and e and m are the electronic charge and mass.

Type (*b*)

For type (*b*) we assume that the maximum electron density is the same in all vertical sections, but the semi-thickness y_m varies linearly in the horizontal direction. We therefore write $y_m = y_{m0}(1+\epsilon X)$, ϵ being a constant. Assuming as before that N at a given x is parabolic in y, we have

$$\mu^2 = 1 - k^2\left\{\frac{2Y}{1+\epsilon X} - \frac{Y^2}{(1+\epsilon X)^2}\right\}. \qquad \ldots\ldots(2)$$

Practical values of λ and ϵ may be estimated from published ionospheric data.

§3. DIFFERENTIAL EQUATION OF THE RAY. GRAPHICAL SOLUTIONS

The differential equation of the path of a ray in a medium in which the refractive index varies in two dimensions can be found by applying the calculus of variations to Fermat's principle $\delta\int\mu ds = 0$ (or alternatively by imagining the medium between each two successive surfaces of refractive index $\mu - \delta\mu$ and $\mu + \delta\mu$ replaced by a medium of refractive index μ, and applying Snell's law to the discontinuous medium thus formed). The differential equation obtained may be written in many ways. For graphical step-by-step tracing of a ray a convenient form is $d\theta = -\tan\phi\, d(\log\mu)$ where θ is the angle between the vertical and the direction of the ray, and ϕ is the angle between the direction of the ray and the direction of grad μ. This equation may also be written as $d\theta = -\sin\phi\,|\operatorname{grad}(\log\mu)|\,ds$ where ds is the element of path length. This latter form is convenient when ϕ approaches $\pi/2$.

§4. ANALYTICAL SOLUTIONS OF THE DIFFERENTIAL EQUATIONS

In addition to the forms already given, the differential equation of the path may be written in either of the forms

$$\frac{d}{dS}(\mu\sin\theta) = \frac{\partial\mu}{\partial X}$$

$$\frac{d}{dX}(\mu^2\sin^2\theta) = \frac{\partial\mu^2}{\partial X}$$

where θ is defined as before and is equal to $\sin^{-1} dX/dS$, where dS is the element of path length measured with y_{m0} as unit. The equation when written in the second form is seen to be immediately integrable for an ionosphere in which $\partial\mu^2/\partial X$ is independent of Y and Rawer (1951) has investigated an ionosphere of this type. Our types (*a*) and (*b*) do not have this property, and since it imposes an arbitrary relationship between the rate of horizontal variation of the central density and the rate of horizontal variation of the width, we have thought it better, in the present paper, to develop methods which, though exemplified by application to the types (*a*) and (*b*), are of general applicability. Denoting the angle of incidence by i, we may integrate the equations along the path from the point of incidence, where $\mu = 1$, thus obtaining

$$\mu\sin\theta - \sin i = \int\frac{\partial\mu}{\partial X}dS = \int\frac{1}{2\mu\sin\theta}\frac{\partial\mu^2}{\partial X}dX \qquad \ldots\ldots(3\,a)$$

$$\mu^2\sin^2\theta - \sin^2 i = \int\frac{\partial\mu^2}{\partial X}dX. \qquad \ldots\ldots(3\,b)$$

These equations represent the path of the ray, but the integrals on the right-hand sides, being taken along the path, require a prior knowledge of it.

(i) *Approximate Solutions for Gradient of Type (a)*

If λ is sufficiently small we may regard the effect of the horizontal gradient as that of a small perturbation; hence the right-hand sides of (3) can be regarded as small correction terms. The path of the ray will not differ greatly from the path obtained by putting $\lambda=0$. Thus for a first approximate solution of either (3 *a*) or (3 *b*) we have to evaluate $(\partial\mu^2/\partial X)\,dX$ along this path (for only a second order error will be introduced by putting $\mu\sin\theta=\sin i$ under the integral sign in (3 *a*)). To evaluate this integral for gradient of type (*a*) we note that, from (1), $\partial\mu^2/\partial X=-\lambda k^2(2Y-Y^2)$ and that the equation of the path for $\lambda=0$ is readily shown to be

$$Y=1-(1-C^2/k^2)^{1/2}\cosh\left(\frac{kX}{S}-\beta\right) \qquad \ldots\ldots(4)$$

where $C=\cos i$, $S=\sin i$, $\tanh\beta=C/k$. Evaluating the integral we find that, to the degree of accuracy considered, (3 *a*) and (3 *b*) are

$$\mu\sin\theta-S=-\frac{\lambda}{2S}\left\{\left(\frac{k^2+C^2}{2}\right)X-\frac{S(k^2-C^2)}{4k}\sinh 2\left(\frac{kX}{S}-\beta\right)-\frac{CS}{2}\right\} \ldots(4\,a)$$

$$\mu^2\sin^2\theta-S^2=-\lambda\left\{\left(\frac{k^2+C^2}{2}\right)X-\frac{S(k^2-C^2)}{4k}\sinh 2\left(\frac{kX}{S}-\beta\right)-\frac{CS}{2}\right\}. \ldots(4\,b)$$

The simple relationship between these approximate forms of (3 *a*) and (3 *b*) is due, of course, to $\mu\sin\theta+S$ being approximately $2S$, and the equations (4 *a*) and (4 *b*) will naturally give slightly different values for $\mu\sin\theta$. To obtain approximate expressions for the angle of emergence e, we have to insert for X the value $2\,\beta\,S/k$, since this is the coordinate of the point of emergence of the ray, for $\lambda=0$. Thus, since $\mu=1$ at emergence, these approximate expressions are

$$\sin e-\sin i=-\tfrac{1}{2}\lambda k\{(1+C^2/k^2)\beta-C/k\} \qquad \ldots\ldots(4\,c)$$

$$\sin^2 e-\sin^2 i=-\lambda kS\{(1+C^2/k^2)\beta-C/k\}. \qquad \ldots\ldots(4\,d)$$

(ii) *Approximate Solutions for Gradient of Type (b)*

Here we regard ϵ as the small perturbing quantity, and we note that, from (2), $\partial\mu^2/\partial X=2\epsilon k^2Y(1-Y)$ to the first order in ϵ. We then deduce the approximate expressions

$$\mu\sin\theta-S=\frac{\epsilon}{2S}\left\{2S(k^2-C^2)^{1/2}\sinh\left(\frac{kX}{S}-\beta\right)\right.$$
$$\left.-\frac{S}{2k}(k^2-C^2)\sinh 2\left(\frac{kX}{S}-\beta\right)-(k^2-C^2)X+CS\right\} \qquad \ldots\ldots(5\,a)$$

$$\mu^2\sin^2\theta-S^2=\epsilon\left\{2S(k^2-C^2)^{1/2}\sinh\left(\frac{kX}{S}-\beta\right)\right.$$
$$\left.-\frac{S}{2k}(k^2-C^2)\sinh 2\left(\frac{kX}{S}-\beta\right)-(k^2-C^2)X+CS\right\}. \ldots\ldots(5\,b)$$

The expressions for the angle of emergence are, for gradients of this type,

$$\sin e-\sin i=\epsilon k\{C/k-(1-C^2/k^2)\beta\} \qquad \ldots\ldots(5\,c)$$

$$\sin^2 e-\sin^2 i=2\epsilon kS\{C/k-(1-C^2/k^2)\beta\}. \qquad \ldots\ldots(5\,d)$$

§5. HORIZONTAL RANGE IN THE IONIZED LAYER

Consider first type (*a*). We will write $\mu_0^2 = 1 - k^2(2Y - Y^2)$ and $\gamma = k^2X(2Y - Y^2)$; thus $\mu^2 = \mu_0^2 - \lambda\gamma$. If, then, we write (4 *b*) as $\mu^2 \sin^2\theta - S^2 = -\lambda F$ we find (remembering that (dY/dX) is $\cot\theta$) that this equation may be written

$$\int dX = \int \frac{S\,dY}{\{\mu_0^2 - S^2 + \lambda(F\mu_0^2 - \gamma S^2)/S^2\}^{1/2}}.$$

It will be noted that equations (4 *a*) and (4 *b*) yield the same expression for $\mu^2 \sin^2\theta - S^2$ if terms in λ^2 are neglected.

The denominator is to be thought of as a function of Y (the expression $F\mu_0^2 - \gamma S^2$ being given its value, as a function of Y, on the path for $\lambda = 0$). The greatest value of Y reached by the ray is that at which the slope is zero, i.e. at which the quantity under the root vanishes. Provided the end point of the integral is not in this turning region we may expand the integral in the form

$$\int dX = \int \frac{S\,dY}{(\mu_0^2 - S^2)^{1/2}} - \lambda \int \frac{F\mu_0^2 - \gamma S^2}{2S(\mu_0^2 - S^2)^{3/2}}\,dY.$$

If the end point is in the descending part of the ray, the path of integration goes through a region where the expansion is not valid; but by thinking of Y as extended to complex values, we may deform the path into one throughout which the expansion holds. Thus if Δ is the X displacement of the ray (relative to the ray with $\lambda = 0$) at a given Y, not near the top of the trajectory, we have,

$$\Delta = -\lambda \int \frac{F\mu_0^2 - \gamma S^2}{2S(\mu_0^2 - S^2)^{3/2}}\,dY.$$

For the actual evaluation of Δ it is convenient to use the variable dX so that we have to evaluate, along the corresponding path, the expression

$$-\lambda \int \frac{F\mu_0^2 - \gamma S^2}{2S^2(\mu_0^2 - S^2)}\,dX.$$

The result is, for the X displacement Δ at the point of emergence,

$$\Delta = -\frac{\lambda}{2k}\left\{(2 + k^2 - C^2)\frac{\beta^2}{k} - \frac{\beta}{C} + \frac{kS^2}{k^2 - C^2}\right\}. \qquad \ldots\ldots(6\,a)$$

An exactly similar calculation can be carried out for a gradient of type (*b*), working to the first order in ϵ. The result is

$$\Delta = \frac{\epsilon}{k}\left\{(2 - k^2 - C^2)\frac{\beta^2}{k} + \frac{\beta}{C} - \frac{kS^2}{k^2 - C^2}\right\}. \qquad \ldots\ldots(6\,b)$$

§6. CRITICAL PENETRATION FREQUENCY AT OBLIQUE INCIDENCE

For radio transmission calculations it is important to know the greatest frequency $f_{\max}$ that can be reflected from an ionospheric layer at a given angle of incidence i. When λ (or ϵ) is zero $f_{\max}$ is $f_c \sec i$, where f_c is the critical penetration frequency at normal incidence. The existence of a horizontal gradient of ionization modifies this relation.

For type (*a*), as the frequency approaches $f_{\max}$ the path of the ray becomes asymptotic to the line $Y = 1$. Along this line, θ is $\pi/2$, and μ^2 is $1 - k^2(1 + \lambda X)$. Hence if k has the value corresponding to frequency $f_{\max}$ we have from 3 (*b*), on giving the upper limit a sufficiently large value, say L,

$$1 - k^2(1 + \lambda L) - S^2 = -\lambda k^2 \int_0^L (2Y - Y^2)\,dX. \qquad \ldots\ldots(5)$$

We may take the integral along what would be the critical path if λ were zero, viz. the curve $Y = 1 - \exp(-X/T)$ where $T = \tan i$. The equation (5) is thus

$$C^2 - k^2(1 + \lambda L) = -\lambda k^2\{L + \tfrac{1}{2}T \exp(-2L/T) - \tfrac{1}{2}T\}.$$

Since L is to be given a very great value, the term in $\exp(-2L/T)$ is negligible, and we have $k = C/(1 + \frac{1}{2}\lambda T)^{1/2}$ and hence

$$f_{\text{max}} = (1 + \tfrac{1}{2}\lambda T)^{1/2} f_c \sec i. \qquad \ldots\ldots(7\,a)$$

For type (*b*) the ray becomes asymptotic to the line $Y = 1 + \epsilon X$. However, to the first order $\sin^2\theta$ is still unity, and we therefore have, if L is sufficiently great,

$$1 - k^2 - S^2 = 2\epsilon k^2 \int_0^L Y(1 - Y)\,dX.$$

We integrate the right-hand side as before, and obtain $k = C/(1 - \epsilon T)^{1/2}$ so that in this case

$$f_{\text{max}} = (1 + \epsilon T)^{1/2} f_o \sec i. \qquad \ldots\ldots(7\,b)$$

It will be noted that (7 *a*) and (7 *b*) are identical if we replace ϵ by $\lambda/2$.

§7. APPLICATION OF THE ANALYSIS TO THE CASE OF A CURVED IONOSPHERE

The above calculations have dealt with the case of a plane ionosphere. In considering transmission over long distances it is necessary to include the effect of the curvature of the earth and ionosphere.

We use polar coordinates (r, θ), r being measured from the centre of the earth, and a positive θ corresponding to a positive X in the plane case. We will write R = radius of lower side of the ionized layer; $y = r - R$; y_{m0} = semi-thickness of the layer at $\theta = 0$; $y/y_{m0} = Y$; and $R/y_{m0} = \rho$.

The refractive index corresponding to type (*a*) is given by

$$\mu^2 = 1 - k^2(2Y - Y^2)(1 + \lambda\rho\theta); \quad \mu_0{}^2 = 1 - k^2(2Y - Y^2)$$

and that corresponding to type (*b*) by

$$\mu^2 = 1 - k^2\left\{\frac{2Y}{1 + \epsilon\rho\theta} - \frac{Y^2}{(1 + \epsilon\rho\theta)^2}\right\}.$$

Fermat's principle may be expressed in various forms, those analogous to (3 *a*) and (3 *b*) being

$$\mu r \sin\psi - RS = \int \frac{\partial\mu}{\partial\theta}\,ds = \int \frac{r^2}{2\mu r \sin\psi}\frac{\partial\mu^2}{\partial\theta}\,d\theta \qquad \ldots\ldots(8\,a)$$

$$\mu^2 r^2 \sin^2\psi - R^2S^2 = \int r^2 \frac{\partial\mu^2}{\partial\theta}\,d\theta. \qquad \ldots\ldots(8\,b)$$

Here $\psi = \tan^{-1}(r\,d\theta/dr)$ = the angle between the direction of the path and the local vertical. As before the solution will not be exact unless the integral is taken along the actual path. Whether we use (8 *a*) or (8 *b*) the first approximation requires us to evaluate $r^2(\partial\mu^2/\partial\theta)\,d\theta$ along the path for $\lambda = 0$; (replacing $\mu r \sin\psi$ by RS under the integral only introduces a second order error). This path is, by the definition of ψ,

$$\frac{d\theta}{dr} = \frac{RS}{r(\mu_0{}^2 r^2 - R^2S^2)^{1/2}}.$$

We may write this approximately, remembering that Y/ρ is small, as

$$k\rho\,d\theta = S\,dY/\{Y^2 - 2Y(1 - \sigma) + q^2\}^{1/2}$$

where $q = C/k$ and $\sigma = S^2/\rho k^2$; the solution of this is

$$Y = 1 - \sigma - \{(1-\sigma)^2 - q^2\}^{1/2} \cosh\left(\frac{k\rho\theta}{S} - \beta_1\right) \qquad \ldots\ldots(9)$$

where $\tanh \beta_1 = C/k(1-\sigma)$. In evaluating $\int r^2(\partial\mu^2/\partial\theta)\,d\theta$ along this path we may replace r^2 by R^2; and for the angle of emergency e we obtain, for type (*a*),

$$\sin e - S = -\tfrac{1}{2}\lambda k\{(1+q^2+2\sigma)\beta_1 - (1+3\sigma)C/k\} \qquad \ldots\ldots(10\,a)$$

and for type (*b*)

$$\sin e - S = \epsilon k\{C(1-3\sigma)/k - \beta_1(1-q^2-4\sigma)\}. \qquad \ldots\ldots(10\,b)$$

We may also perform a calculation similar to that of § 5, to find the linear displacement of the point of emergence. If, for a gradient of type (*a*) we write (8 *b*) as $\mu^2 r^2 \sin^2\psi - R^2 S^2 = FR^2$ we find that to obtain Δ we have to integrate the expression

$$-\frac{\lambda\rho}{2S^2}\int \frac{F\mu_0^{\,2} - \gamma S^2}{\mu_0^{\,2} - S^2(1-2Y/\rho)}\,d\theta$$

along the path (9). The result is

$$\Delta = \frac{\lambda}{2k}\left[\{2 + k^2(1+2\sigma) - C^2\}\beta_1^{\,2} - \frac{\beta_1}{C}\left(3 - 2C^2\frac{k^2+C^2}{k^2-C^2}\right) + k\,\frac{S^2 + 2\sigma(1-2C^2)}{k^2 - C^2 - 2\sigma k^2}\right].$$

§ 8. LIMITATIONS OF ANALYSIS

The graphical method outlined in § 3 is theoretically capable of any desired accuracy, but it is laborious. The approximate analytical formulae are easy to use but again may be inaccurate when the horizontal gradient of ionization is very large.

An examination of published ionospheric data indicates that in practice the magnitudes of the perturbation terms λ and ϵ are seldom likely to exceed 0·5 and 0·05 respectively, and that typical average values would be 0·3 and 0·03 in winter, and 0·02 and 0·01 in summer. We now consider briefly the accuracy and limitations of the foregoing analysis.

(i) *Gradient Type* (*a*)

Formulae for angle of emergence e

For large λ it is possible to obtain a refined value of e by making an approximate graphical tracing of the path and applying eqn. (3 *b*), evaluating the right-hand side by numerical integration. This method shows that (4 *c*) is a better approximation than (4 *d*). Thus for $\lambda = +\frac{1}{3}$, $f = f_c$, (4 *c*) yields values of e which are in error by less than a degree for i greater than 35°; for $\lambda = \frac{1}{10}$ this accuracy is maintained for i greater than about 20°. It is clear that as C/k approaches unity, the ray penetrates the ionized layer more deeply and encounters greater values of the horizontal gradient: the distortion of the path from its unperturbed shape is thus greater and the accuracy of the approximation diminishes.

The distortion produced in the ray path is greater when λ is negative than when λ is positive, and the errors in using (4 *c*) will be correspondingly greater. For the same accuracy, a positive λ of $\frac{1}{3}$ corresponds to a negative λ of $\frac{1}{10}$. (Note that for f greater than about $2f_c$, the critical penetration formula gives a check point $e = \pi/2$.)

Formula for Δ

The errors in using eqn. (6*a*) can be estimated roughly by tracing the ray paths graphically, and for λ positive appear to be of the order $100\lambda\%$. However, for λ negative and for angles near to the critical penetration angle, Δ becomes much larger than the values given by (6*a*).

Formula for $f_{\max}$

It is difficult to estimate the accuracy of eqn. (7*a*), but the error in $(f_{\text{perturbed}} - f_{\text{unperturbed}})$ should not be more than $100\lambda\%$, since we have only neglected terms of the second and higher order in λ.

(ii) *Gradient Type* (*b*)

ϵ is always so small that all the formulae derived should be accurate as regards deviations from the unperturbed state to within, at most, 5%.

(iii) *Magnitude of Ionospheric Curvature Effect*

Calculation shows that in all but extreme cases the effect of ionospheric curvature will be small and of less importance than other factors likely to be involved in a practical application of the analysis. For example, for gradient type (*a*), the values of *e* calculated for the curved ionosphere case are seldom likely to differ by more than 3% from those calculated for the plane ionosphere case. In view of other limitations affecting the practical application of an analysis of this kind, a detailed study of curvature effects is not considered justifiable.

§ 9. DIAGRAMS

Figures 1 to 5 have been drawn to illustrate one or two typical cases of propagation in the presence of both horizontal and vertical ionization gradients.

Figure 1 shows a ray path traced by means of the graphical method described in § 3 for representative type (*a*) conditions. It will be noted that the ray is

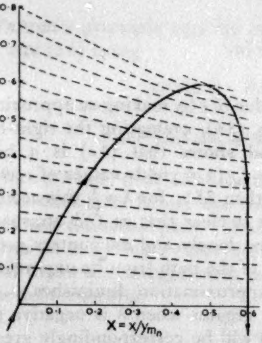


Fig. 1. Typical ray path in a layer with horizontal and vertical gradients type (*a*) $\lambda = \frac{1}{3}, f = f_c$.

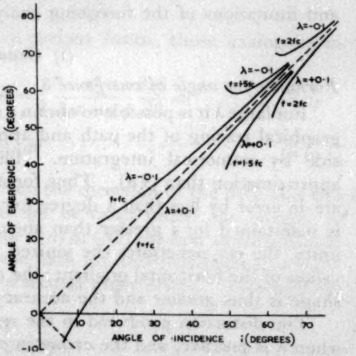


Fig. 2. Angles of incidence and emergence gradient type (*a*) $\lambda = \pm 0{\cdot}1$.

incident at an angle of about 22° and emerges in an approximately vertical direction.

Figure 2 shows angle of incidence and angle of emergence for positive and negative gradients of type (*a*) for $\lambda = \pm 0{\cdot}1$. It will be seen that for $f = f_0$ the ray incident at about 7° is returned in a direction parallel to itself. With $f = f_c$ for both positive and negative values of λ, the angles of emergence of rays incident nearly vertically become very large.

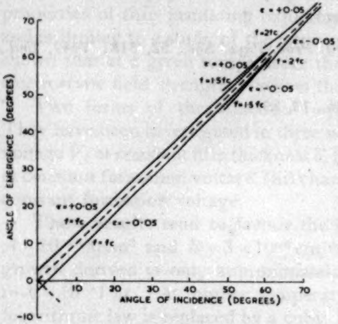


Fig. 3. Angles of incidence and emergence gradient type (*b*) $\epsilon = \pm 0{\cdot}05$.

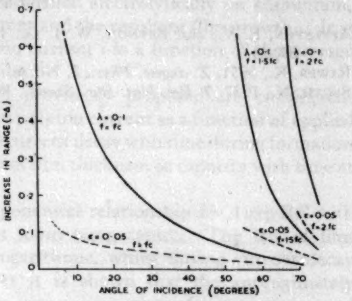


Fig. 4. Increase in horizontal range $(-\Delta)$. Gradient type (*a*) $\lambda = 0{\cdot}1$, gradient type (*b*) $\epsilon = 0{\cdot}05$.

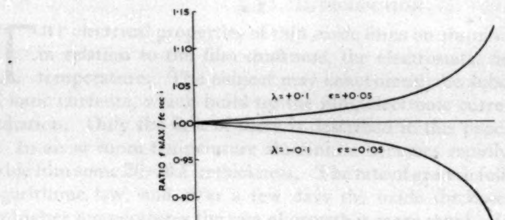


Fig. 5. Fractional change in maximum penetration frequency.

Figure 3 shows similar curves of *i* against *e* for gradient type (*b*) with $\epsilon = \pm 0{\cdot}05$.

Figure 4 shows the increase in the horizontal range in the layer due to a horizontal gradient of ionization for types (*a*) and (*b*) (see § 5).

It will be noted that the effect of horizontal ionization gradient is most serious near the angles corresponding to penetration.

Figure 5 shows the ratio of the penetration frequency with and without a horizontal gradient, i.e. $f_{\max}/f_c \sec i$ (see § 6) for typical positive and negative values of λ and ϵ.

ACKNOWLEDGMENTS

I should like to express my thanks to Dr. W. J. G. Beynon for suggesting the problem and for advice, to Dr. P. M. Davidson for assistance in the mathematics, and to Professor F. Llewellyn Jones for his interest.

This work was carried out as part of the programme of the Radio Research Board of the Department of Scientific and Industrial Research and this paper is published by permission of the Director of Radio Research.

REFERENCES

APPLETON, E. V., and BEYNON, W. J. G., 1940, *Proc. Phys. Soc.*, **52,** 518; 1947, *Ibid.*, **59,** 58.
RAWER, K., 1951, *Z. angew. Phys.*, **3,** No. 6.
SMITH, N., 1937, *J. Res. Nat. Bur. Stand., Wash.*, **19,** 89.

Ionic Current and Film Growth of Thin Oxide Layers on Aluminium

By A. CHARLESBY
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

MS. received 19th March 1952, *and in final form 8th January* 1953

Abstract. This paper describes part of a series of investigations into the electrical properties of thin insulating oxide layers formed electrolytically on aluminium, and is limited to a study of the ionic current and the resultant film growth. It is shown that at a given temperature the ion current i is a function of the average electrostatic field strength F across the film.

Two forms of the function $i=f(F)$ previously proposed are considered. They have been investigated in three ways: (i) ion current as a function of applied voltage V_f at constant film thickness δ, (ii) current decay with time during formation at constant formation voltage, (iii) change of film thickness or capacity with time at constant formation voltage.

These results tend to favour the exponential relationship $i=A \exp BF$ with $A=10^{-18}$ A/cm^2 and $B=3\times10^{-6}$ cm/v at room temperature. The law of film growth derived is only approximately logarithmic, whilst during current decay $i \sim 6\times10^{-5} V_f/t$. At higher temperatures it is shown that the approximately logarithmic law is replaced by a cubic law over a short range of temperatures, and subsequently by a quadratic law. The application of these results to oxidation in air is discussed.

§ 1. Introduction

The electrical properties of thin oxide films on aluminium have been studied in relation to the film thickness, the electrostatic field across it, and the temperature. The subject may conveniently be subdivided into the study of ionic currents, which build up the film, electronic currents, and the effects of radiation. Only the first of these is described in this paper.

In air at room temperature aluminium becomes rapidly coated with a stable oxide film some 20–30 Å in thickness. The rate of growth follows an approximately logarithmic law, and after a few days the oxide thickness is almost constant. At higher temperatures the rate of growth is more rapid. In suitable electrolytes it is possible to form oxide films of considerably greater thickness than are obtainable in air at room temperature. Two types of oxide film may be formed, depending on the type of electrolyte used: (*a*) A so-called amorphous film, produced, for example, when a solution of oxalic acid is used as electrolyte. Relatively thick films may be formed at moderate voltages (<100 v), the thickness of oxide produced depending primarily on the total charge passed, and not on the applied voltage. (*b*) An insulating film, produced, for example, in dilute ammonium borate, where thickness depends on the applied voltage. It is rarely possible to exceed 400–600 v, or film thicknesses of a few thousand ångströms at room temperature. Evidence has been produced (Dekker and Van Geel 1947) to show that the difference between these two types of oxide is due to the presence of pores in the former. In this paper only the latter or insulating type of oxide layer will be considered.

In the literature reference is often made to the formation current (generally of the order of mA/cm^2) which flows during the initial process of formation, and to the leakage current (of the order of μA/cm^2) which persists after formation is complete and whose cause has not yet been studied. In this paper we prefer to consider a *current of ions* i_+ (which builds up the film) and *an electron current* i_- which flows through the film with no obvious permanent effect on the latter. These two currents are both assumed to occur simultaneously, but in widely differing proportions. The formation efficiency η may then be defined as $i_+/(i_-+i_+)$. Provided i_+ is not too small (e.g. 1 mA/cm^2 during formation) η is close to unity. Only the behaviour of the ionic current is considered in this paper.

In the following sections the thickness of the insulating oxide is represented by δ and the potential applied across it by V. When the current through the cell is low it is often sufficient to take V as the potential applied externally across the cell. The average electrostatic field F across the film equals V/δ, although space charges may cause local variations. Suffix 'f' is used to indicate the values of V, i_+ and F during formation; q represents the charge per unit area of ions passed during film formation and equals η times the total charge passed per cm^2. For a uniform film q is proportional to the film thickness δ, so that we may put $q=r\delta$, where r is a constant at a given temperature.

§2. MEASUREMENT OF FILM THICKNESS

When aluminium or certain other metals (e.g. tantalum) are anodized in suitable electrolytes to form insulating oxide layers, a large capacity is formed between the metal and the electrolyte. This capacity is assumed to be due to the oxide layer, which acts as the dielectric of a condenser, the plates of which are the metal and the electrolyte. Measurement of the capacity will then give information on film thickness.

In its simple form the oxide dielectric theory fails to account for a number of phenomena which are observed in electrolytic condensers. These include (*a*) the rectification properties of the film, (*b*) the very complex relationship between current through the film and the applied potential difference under various conditions of film formation, (*c*) a number of time-dependent properties, such as film ageing, current hysteresis, stability or otherwise of the capacity, (*d*) the effect of ultra-violet radiation. A discussion of these phenomena must be deferred until the electronic current is studied.

An enormous amount of experimental data on these matters is available in the literature, but the theoretical explanations are often inadequate. In addition to our own results we have drawn largely on the fund of detailed experimental data obtained by Guntherschulze and his school (Betz, Grunert and Keller) at Dresden.

If an insulating oxide film is formed by anodizing aluminium at constant current, the formation voltage V_f is found to increase uniformly with time (fig. 1, stage 1) up to about 500 v, when breakdown occurs. At the same time the total capacity C varies inversely with time, so that if we assume $V_f=0$ for a clean metal surface, the product V_fC remains constant, independent of film thickness over the range studied. This is shown in fig. 2, where a logarithmic plot of V_f against C gives a linear curve of slope -1. Since $C=\epsilon/4\pi\delta$ in e.s.u. per unit area, where ϵ is the dielectric constant, it follows that $\epsilon V_f/\delta$ is independent of film thickness δ. The linear growth of both V_f and $1/C$ with time implies that the surface oxide is uniform in structure, so that ϵ is constant. Thus at a given temperature the average formation field F_f ($=V_f/\delta$) across the film, corresponding to a given ion current, is independent of the film thickness. The dependence of V_f/δ on temperature and formation current is not very marked, and is often ignored. This dependence probably explains the discrepancy between the results of Guntherschulze and Betz, and those of Burgers, Claassen and Zernike (1932), to which the latter have drawn attention. A number of attempts have been made to measure the ratio δ/V_f, either by a determination of ϵ, which is not necessarily the same as for the bulk material, or by direct weighing, assuming the film density to be that of corundum. The values obtained include 12·5 Å/v (Dekker and Van Geel 1947), 10·6 to 13·4 Å/v (Guntherschulze and Betz 1932 a, b, c, 1934 a, b) and 11·8 (Deryagin and Friedland 1948). Other metals having similar insulating oxide layers gave values for δ/V_f of 9 A/v (tantalum), 12·5 (zirconium) and 9·6 (tungsten). The ion charge q per unit area required to build a uniform film of thickness δ, formed at voltage V_f, is found to be given approximately by the equation

$$q \sim 2 \times 10^4 \delta \sim 2 \times 10^{-3} V_f \text{ coulombs/cm}^2 \qquad \ldots\ldots(1)$$

at room temperature.

The film thickness δ can either be obtained by a measurement of the capacity per unit area, or of the formation voltage, or of the ion charge passed. Because of the uncertain values of the dielectric constant ϵ, the density ρ and the surface area A none of these measurements is very reliable.

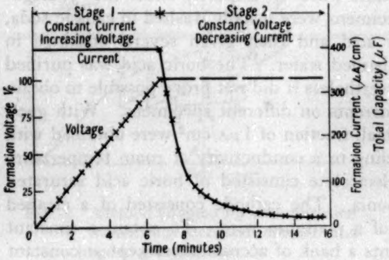


Fig. 1. Aluminium oxide film formation.

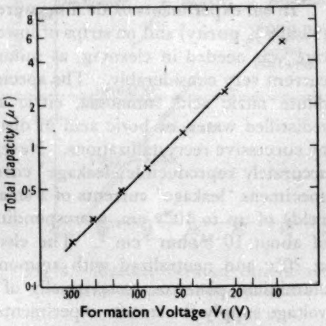


Fig. 2. Capacity and formation voltage. CV_f=constant at constant current.

§3. Ion Current as a Function of Field

At constant formation current i_f the field F_f is independent of film thickness over a wide range. If the formation voltage V_f is maintained constant it has been shown by the writer that during the subsequent decay in formation current (due to increasing film thickness) i_f is a function of the ratio V_f/δ, i.e. of the field F_f, and is independent of film thickness. Under conditions where the electron current is negligible compared with the formation current we may write $i_+ = f(F_f)$ at a given temperature. Experiments show that a large change in i_+ results from only a slight change in F_f; the product CV_f increases only slightly if the formation current is increased tenfold.

Guntherschulze and Betz (1934 b) studied this relationship by measurements of capacity and formation rate. They found that the following formulae fit their experimental data equally well:

I. $i_+ = A\exp(BF)$; $A = 3{\cdot}623 \times 10^{-23}$ A/cm², $B = 4{\cdot}246 \times 10^{-6}$ cm/v,

II. $i_+ = AF^2\exp(-B/F)$; $A = 6{\cdot}61$ A/v, $B = 4{\cdot}324 \times 10^8$ v/cm.

The values of F studied vary over a short range only, $9{\cdot}3 \times 10^6$ to $11{\cdot}2 \times 10^6$ v/cm, so that the exponent is about ±40. Formula I agrees with the theoretical formula deduced by Mott, while formula II is similar to the strong field emission formula due to Fowler and Nordheim (1928), and has been suggested experimentally by Van Geel (1931). With the method used by Guntherschulze and Betz it is not possible to distinguish between these formulae nor can the range be extended, owing to the very high current densities involved at higher fields and to the intrusion of very appreciable electron currents at the lower fields. Alternative methods have therefore been adopted in the present work. The various formulae derived may be of value in other connections. In the first method the variation of i_+ with V_f at constant δ is determined in an indirect manner. In the second method the decay of formation current with time at constant V_f is studied, and in the third the growth of film thickness with time at constant V_f is given. Errors may arise from uncertainty in the exact surface area of the specimen, and from the intrusion of the electron current at low current densities, but do not affect all these methods.

In our experiments oxide films were formed on 'spec-pure' rods of aluminium (99·999% purity) and on strips of lower purity (99·9%). Before anodizing, great care was needed in cleaning, as minute traces of impurity affect the 'leakage' current very considerably. The specimens were usually washed in caustic soda, dilute nitric acid, ammonia, citric acid and then given several soakings in redistilled water, or boric acid in distilled water. The boric acid was purified by successive recrystallizations. Nevertheless it did not prove possible to obtain accurately reproducible 'leakage' currents on different specimens. With good specimens 'leakage' currents of a small fraction of 1 μA/cm² were obtained with fields of up to 10^7 v/cm, corresponding to a conductivity at room temperature of about 10^{-14} ohm^{-1} cm^{-1}. The electrolyte consisted of boric acid saturated at 20°C and neutralized with ammonia. The cathode consisted of a cleaned aluminium plate, or more usually of a platinum wire. To obtain a constant voltage supply for certain experiments a bank of accumulators kept at constant temperature and trickle-charged from the mains was used. Experiments were carried out at a temperature of 20° ± 1°C.

Two methods of anodizing the specimen may be adopted: (*a*) A constant voltage is applied directly between cathode and anode. The current, initially very high, drops off rapidly until after a few hours a fairly constant value is reached. (*b*) A resistance in series with the cell is varied to maintain the current constant. During this process the potential across the cell increases (stage 1 of fig. 1). Once the resistance has been eliminated, the voltage becomes constant, and the current decreases (stage 2). This method is illustrated in fig. 1, which gives the (V, t) and (i, t) curves for an aluminium sheet specimen.

There appears to be little difference between films formed in these two ways, but method (*b*) gives information on the variation with film thickness of voltage necessary to pass a given current.

Method I: Relationship between Current and Voltage at Constant Film Thickness

During the process of measurement the film thickness is constantly increasing owing to the flow of ion current. To circumvent this difficulty use is made of the fact that, at constant formation current, the formation voltage increases uniformly with time. During the formation process the current is reduced (fig. 3) in a series of steps, each lasting 45 (or 60) seconds, to allow the current to steady and the voltage to resume its upward trend. When the lowest convenient formation current is reached the current is again increased in a similar fashion, the whole process being symmetrical about AA′ (corresponding to some film thickness δ_m). Formation current and voltage vary with time in the manner shown in the figure.

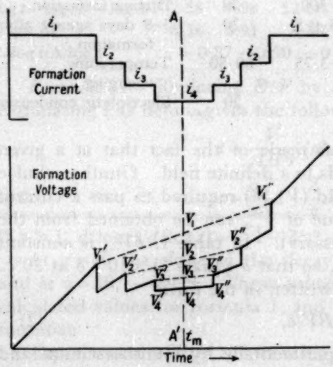


Fig. 3. Method adopted for determination of relationship between current and voltage in growing film.

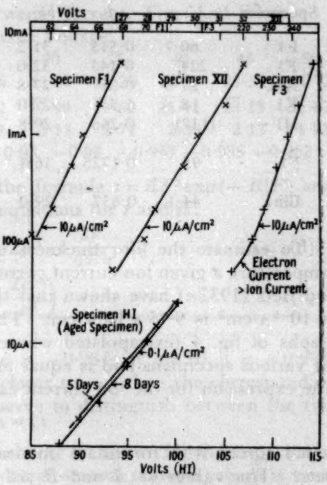


Fig. 4. Relation between current and voltage at constant film thickness.

It can be readily shown that the average value $\frac{1}{2}(V_1' + V_1'') = V_1$ is the voltage which would be required to pass current i_1 through the film at the thickness δ_m. Similarly $\frac{1}{2}(V_2' + V_2'') = V_2$ would be required to pass a current i_2 at thickness δ_m. Plots of $V_n = \frac{1}{2}(V_n' + V_n'')$ against $\log i_n$ are shown in fig. 4. It is seen that the plot is linear except at the lowest current densities, where the electron current is no longer negligible. Over the range considered $i_n = A \exp(bV_n)$, where A, b are constants for any given film. Values of b for different specimens are given in table 1. V_{-5} is the voltage corresponding to a current density of 10^{-5} A/cm².

A similar experiment was carried out with a well-aged film HI (held at a constant potential of 100 v for several days). For this specimen there was no appreciable change in film thickness in the course of the measurements. On the other hand there was a considerable time lag between changes of voltage and of current (of the order of minutes). Even with the very low currents involved (about one thousand times smaller than in the first four specimens) the same value of bV_{-5} is found (V_{-5} must be obtained by extrapolation). Two sets of readings were taken, one after 5 days, the other after 8 days formation at 100 volts. A single experiment carried out at a higher temperature gave a smaller value of bV_{-5} (as was to be expected from theoretical considerations). Owing to the variation of film thickness and the temperature range, no quantitative conclusions can be drawn from this specimen.

Table 1. Relation between Formation Current and Voltage at Constant Film Thickness

Specimen	V_{-5}	b	bV_{-5}	Range (μA/cm^2)	T (°c)	Remarks
F1	60·7	0·515	31·2	7·5–375	20	During formation
F3	224	0·143	32·0	3·7–375	20	During formation
XII	27·8	0·85	27·8	4–400	20	During formation
XI	18	1·5	27·0	10–500	20	During formation
HI	(112)	0·266	29·8	0·05–0·5	20	5–8 days ageing after formation
F1	95	0·1725	16·4	1·5–3·75	60–50	Temperature changing
Elh1	44·1	0·657	29·0		20	Electrolytic condenser

To estimate the film thickness use was made of the fact that at a given temperature a given ion current corresponds to a definite field. Guntherschulze and Betz (1932 a) have shown that the field (V_{-5}/δ) required to pass a current of 10^{-5} A/cm^2 is $9{\cdot}45 \times 10^6$ v/cm. The value of V_{-5} can be obtained from the graphs of fig. 4 (extrapolated where necessary). In table 1, bV_{-5} is constant for various specimens and is equal to 29·5, so that b equals $31 \times 10^{-7}/\delta$ at 20°c. The expression for the ion current can be written in the form

$$i_+ = A \exp BV/\delta, \qquad \ldots\ldots(2)$$

which agrees with formula I obtained experimentally by Guntherschulze and Betz. The values of A and B are however different. In these experiments A equals 10^{-18} to within a factor of 4, and $B = 31 \times 10^{-6} \pm 5\%$, at 20°c.

Over a small range of values of V close to V_1 formula II can be written $\Delta \log i_+ = nB\delta\Delta V/V_1^2$, where n varies by about 1% over the range of values of V studied. Thus formulae I and II represent the observations of fig. 4 equally well. The range of values of V cannot be extended since, at lower current densities, the electron current overshadows the ion current and hysteresis effects appear, while at much higher current densities considerable voltage losses occur, and the film growth is extremely rapid.

Method II: Current Decay at Constant Voltage

Once the maximum voltage of formation is reached, the ion current falls off, rapidly at first, and then more slowly (stage 2, fig. 1). If this decrease is ascribed to the increasing film thickness and resultant drop in field an expression may be derived for the rate of current decay in terms of various possible relationships between current and field.

Formula I. If $i = A\exp(BF)$ and $\delta = rq$, then eliminating i and integrating by parts gives an expression for t:

$$t = \frac{BV}{Ar}\frac{e^{-z}}{z}\left[\frac{1}{z} - \frac{2!}{z^2} + \frac{3!}{z^3} - \dots\right], \qquad \dots\dots(3)$$

where $z = BF$, provided $z \gg 1$. Moreover

$$it = \frac{q}{z}\left[1 - \frac{2!}{z} + \frac{3!}{z^2} - \dots\right]. \qquad \dots\dots(4)$$

Corresponding values of z, i and t are given in table 2; the expansion converges satisfactorily provided $z \gg 1$. The maximum error is about 1·5% for $z = 10$, 0·3% for $z = 15$, and less than 0·1% for $z \geqslant 20$. The slope $\Delta \log i / \Delta \log t$ for the decay curve is also given.

Table 2. Calculated Relationship between Current, Time and Film Thickness at Constant Voltage

$z = BF = BV/\delta$	10	15	20	25	30	35	40	45	50
$\log i - \log A$	4·34	6·51	8·69	10·86	13·03	15·20	17·37	19·54	21·72
$\log t + \log Ar - \log BV$	$\overline{7}$·58	$\overline{9}$·08	$\overline{12}$·67	$\overline{14}$·32	$\overline{17}$·99	$\overline{19}$·69	$\overline{21}$·41	$\overline{23}$·13	$\overline{26}$·87
$it/q \times 10^2$	8·36	5·91	4·56	3·71	3·13	2·71	2·39	2·13	1·93
$\Delta \log i / \Delta \log t$	−0·87	−0·90	−0·92	−0·93	−0·94	−0·947	−0·955	−0·962	

Formula II. Replacing B/F by z in the formula $i = AF^2\exp(-B/F)$ and eliminating i as before gives the following equations for t and it:

$$t = \frac{V}{AB^3r}(z^2 - 2z + 2)e^z - \frac{2V}{AB^3r} \qquad \dots\dots(5)$$

and

$$it = \frac{q}{z}\left(1 - \frac{2}{z} + \frac{2}{z^2} - \frac{2e^{-z}}{z^2}\right). \qquad \dots\dots(6)$$

If $z \gg 1$, $d(\log i)/d(\log t) = -1 + 2/z^2 - 4/z^3$.

At $z = 10$ the slope of the decay curve is −0·984, at $z = 20$ it is −0·995, and at $z = 30$, −0·998. These values are closer to −1 than the corresponding calculated values for formula I, and may serve to distinguish between the two formulae.

Experimental results. Plots of $\log i$ against $\log t$ are shown for three specimens (figs. 5, 6). For low values of t (of the order of several minutes) there is some ambiguity, as the origin of time is unknown. The formulae assume that the measured voltage is applied across the oxide layer, whereas ohmic losses are only inappreciable at a more advanced state of formation. The corrected plot should therefore be of $\log i$ against $\log(t + t_0)$, where t_0 is unknown but small. It is seen in fig. 6 that only one value of t_0 gives a linear plot, and that in any case errors in t_0 do not affect appreciably the slope for $t > 10$ minutes.

For these three specimens the slopes $d(\log i)/d(\log t)$ of the decay current are −0·96 (U), −0·92 (H) and −0·96 to −0·98 (RA1). These values are fairly close to the calculated slope for $z \sim 31$ in formula I (i.e. −0·94), but are appreciably different from that deduced from formula II (−0·998). The results tend to favour formula I, but cannot be taken as conclusive.

The value for it/q in table 2 is seen to vary only very slowly with time. An increase in time by a factor of 100 only decreases it/q by some 10–20% when $z \sim 30$. During current decay $it \simeq 3 \times 10^{-2}\, q$ or $6 \times 10^{-5}\, V_f$ coulomb/cm². The constancy of the product it has been observed in previous experiments but has now been derived theoretically.

Method III: Change in Capacity with Time

During stage 2 of the formation process, when the formation voltage remains constant, the formation current decreases as the film becomes thicker. The rate of increase of film thickness can be calculated in terms of the relationship between ion current and field, and can also be measured on a capacity bridge, so that the relationship can be verified.

In table 2 (formula I) it is seen that there is an approximately linear relation between $\log t$ and z: $\Delta \log t = -m\Delta z = -mBV\Delta(1/\delta) = -mBF\delta\Delta(1/\delta)$, where m can be deduced from table 2, and varies by about 1% for a hundredfold variation in t. Writing δ, $\Delta(1/\delta)$ in terms of C, ΔC, we obtain the equation

$$\Delta \log t = -mBF\Delta C/C. \qquad \ldots\ldots(7)$$

This result gives the increase in capacity and the value of BF independently of the true surface area (which is unknown) or the electron current and ion currents flowing. If $BF \sim 30$, $m = 0{\cdot}46$. The variation of BF with time may be neglected after the first few minutes of formation.

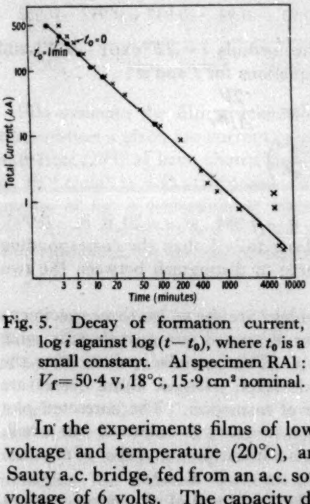


Fig. 5. Decay of formation current, $\log i$ against $\log (t-t_0)$, where t_0 is a small constant. Al specimen RAl: $V_f = 50{\cdot}4$ v, 18°c, 15·9 cm² nominal.

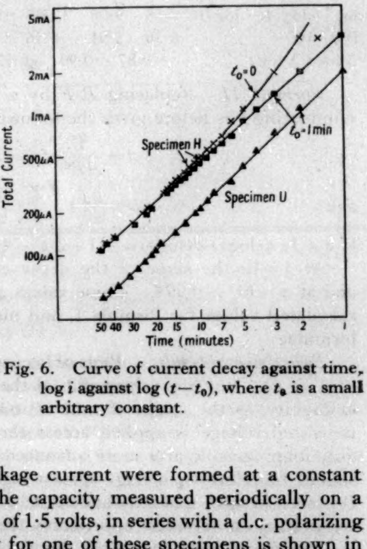


Fig. 6. Curve of current decay against time, $\log i$ against $\log (t-t_0)$, where t_0 is a small arbitrary constant.

In the experiments films of low leakage current were formed at a constant voltage and temperature (20°c), and the capacity measured periodically on a Sauty a.c. bridge, fed from an a.c. source of 1·5 volts, in series with a d.c. polarizing voltage of 6 volts. The capacity decay for one of these specimens is shown in fig. 7. The values of BF deduced for three specimens are given in table 3. They are in fair agreement with those obtained in Method I, although the ion currents

Table 3. Capacity Decay

Specimen	V_f(v)	$\dfrac{-\Delta C}{\Delta \log t}$ (μF)	C (μF)	BF
R2	49	0·1	1·4	31
H10	101·5	0·2	3·7	40
RAl	50·5	0·115	1·9	36

differ by factors of up to 10^4. Figure 8 shows a similar plot for some experimental data given by Guntherschulze and Betz in the case of tantalum metal, for which $B \sim 3{\cdot}5 \times 10^{-6}$.

Electrolytic condensers using an aluminium oxide dielectric are known to suffer a decrease in capacity with time. Explanations given in the literature include the loss of electrolyte and clogging of the porous film by deposits of crystals or colloidal aluminium salts. These hypotheses appear unnecessary since the theory predicts such a drop without further assumptions. Some data for electrolytic condensers, obtained from Coursey and Ray (1939, fig. 17, p. 121), are given in table 4. As these values are taken from a graph and the points are scattered, the results are not very accurate. The values of BF deduced from them are nevertheless of the right magnitude, and indicate that film growth in accordance with the above formula would account for the observed decay in capacity.

Table 4. Change in Capacity with Time (Electrolytic Condensers)

Condenser type	Capacity (μF) Nominal	50 days	500 days	$\dfrac{\Delta C}{\Delta \log t}$	BF
Dry	18	22·8	20·3	2·5	18
Dry	18	21·4	19·4	2·0	23
Wet	16	15·4	14·8	0·6	53

The average value of BF, 30, agrees well with that obtained above, differences between the specimens being possibly due to changes in applied potential in the course of a year, or of temperature. The ion current required to produce this small rate of growth is of the order of 3×10^{-6} mA cm^{-2}, whereas in method I ion currents were of the order of 1 mA cm^{-2}. The agreement illustrates the wide range of validity of the formula for the ion current.

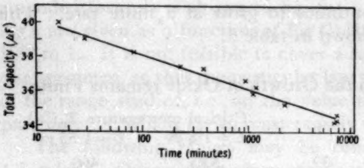


Fig. 7. Capacity decrease with time (aluminium).

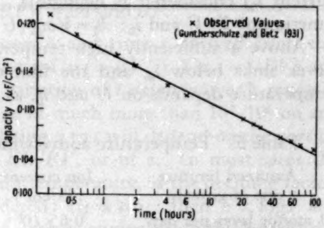


Fig. 8. Capacity decay (tantulum specimen; $V_f = 102{\cdot}9$ v, 0°c).

§ 4. MOTT'S THEORY OF OXIDATION

Mott (1947) assumes that in the formation of an oxide film on aluminium, metal ions leave the metal under the influence of an electrostatic field F, due to an electronic charge on the air interface. This field assists the ions to surmount the potential barrier (of half-width a and height U) at the metal–oxide interface. The formula for ion current flow is as follows:

$$i_+ = N\nu q_1 \exp(-U/\boldsymbol{k}T)\exp(aq_1F/\boldsymbol{k}T) \qquad \ldots\ldots(8)$$

provided that $aq_1F \gg \boldsymbol{k}T$. q_1 is the charge per ion, N the number of metal atoms/cm^2 suitably located for movement into the oxide layer, and ν is the frequency of atomic vibration. A similar type of formula has been deduced by Verwey (1935 a, b, c) on the assumption that the limiting factor is the rate at which ions move from one interstitial position to another, under the influence of the field F.

The Mott formula is of the same type as that found experimentally ($i_+ = A \exp BF$). With our values for A and B, and with the values of N (10^{15}/cm^2), ν (10^{12}/sec) and $q_1(3\boldsymbol{e})$ as chosen by Cabrera and Mott (1948), the dimensions of the barrier limiting ion flow are: $U = 1{\cdot}55$ev, $a = 2{\cdot}5$ Å. The values deduced from the results of Guntherschulze and Betz are higher: $U = 1{\cdot}8$ ev, $a = 3{\cdot}5$ Å.

These values for the barrier dimensions appear too high to substantiate the hypothesis suggested by Verwey. At least part of the oxide film consists of $\gamma\,Al_2O_3$, in which there is one vacant lattice site per eight aluminium ions. Thus there is adequate interatomic distance for ion motion. Ag_2HgI_4 has a similar open structure (3 metallic ions per 4 lattice points) and shows a barrier height of 0·19 ev. In NaCl the barrier height is believed to be 0·90 ev. It would therefore appear that the potential barrier opposing ion flow is located at the metal surface.

§ 5. EFFECT OF TEMPERATURE ON FILM THICKNESS

According to eqns. (2) and (7), the oxide layer continues to grow whenever a field is applied. Nevertheless it is possible to define a limiting field thickness δ_1 corresponding to a given formation voltage and temperature, beyond which the rate of growth is less than some arbitrarily chosen rate (equivalent to an ion current i_0) considered as negligible. Replacing F by V/δ_1 in (8), gives δ_1 as a function of T, V and i_0: $\delta_1 = Vaq_1/(U - n\boldsymbol{k}T)$ with $n = \log(N\nu q_1/i_0)$.

Above a sufficiently high temperature T_c (equal to $U/n\boldsymbol{k}$) the ion current never sinks below i_0, and the film continues to grow at a finite rate. This temperature depends on U and i_0, as shown in table 5.

Table 5. Temperature above which the Growth of Oxide remains Finite

Assumed limiting rate of growth	Ion current (A/cm^2)	n	Critical temperature T_c (°K)	
			$U = 1{\cdot}8$ ev	$U = 1{\cdot}55$ ev
1 atomic layer per day	$0{\cdot}6 \times 10^{-8}$	39	530	460
10 atomic layers per hour	10^{-6}	34	610	530

To check this variation of limiting film thickness with temperature the capacity of oxide layers formed at increasing temperatures was measured. Since

$$C = \frac{\epsilon}{4\pi\delta_1} = \frac{\epsilon U}{4\pi Vaq_1} - \frac{\epsilon nkT}{4\pi Vaq_1} \qquad (n \sim 34,\ 39)$$

a plot of C against T should give a straight line; and at T_c, $C = 0$. In practice it is necessary to extrapolate to $C = 0$, since the formulae used are not valid for thick films, when z or BV/δ_1 is less than about 5.

In the experiments an oxide film was formed by anodizing a specimen for two hours at 36 v at each of a number of increasing temperatures. After each formation the capacity of the film was measured at 20°C. The graph of C against T was linear, giving on extrapolation $C = 0$ at a temperature $T_c \sim 600$°K. Grunert

(1934) carried out a similar set of measurements; his results also showed a linear decrease of C with T, but the critical value of T_c was lower (440°K). Unfortunately Grunert does not state whether all the capacity measurements were made at the same temperature. This is essential since C varies with temperature even for an oxide film of constant thickness. However, it is to be expected that the value of T_c deduced from Grunert's results would be lower, since formation at each temperature lasted for twenty hours, resulting in a smaller final ion current and a larger value of n.

§ 6. OXIDATION AT ROOM AND AT HIGHER TEMPERATURES

The oxidation of aluminium in air at room temperature has been discussed by Cabrera and Mott (1948). The logarithmic law of growth observed and the limiting film thickness in air correspond to anodization in an electrolyte under a formation voltage of about 2 v. Then if $it \sim 6 \times 10^{-5} V_f$, $i \sim 10^{-9}$ A cm^{-2} after one day. This value agrees approximately with the observed rate of growth.

If the film thickness is increased to a marked extent, for example by raising the temperature of oxidation in air above 200°c, the approximate logarithmic relationship between z and t (or δ and t) given on page 323 fails since the integration methods used produce diverging results. This failure of the logarithmic oxidation law occurs for $z<5$. It is therefore necessary to review the basic formula adopted, i.e. I.

The formula $i_+ = A \exp(BF)$ based on Mott's theory considers the flow of metallic ions from the metal into the oxide under the influence of the electrostatic field. When this field is small, the flow of ions in the opposite direction is no longer negligible. The net flow is then

$$i_+ = A\{\exp BF - \exp(-BF)\} = 2A \sinh BF.$$

The relationship between the film thickness δ and time t derived from this amended formula is shown in fig. 9. Plots of z $(=BV/\delta)$, $1/z$ $(=\delta/BV)$ and $\log z$ are given as a function of $\log t$; the variation of t extends over a range of 10^{30} to 1. It is not feasible to cover a range of much more than 10^3–10^5 on any one specimen, so that the particular law relating z to t will depend on the portion of the range studied, i.e. on the value of tAr/BV, or of z. In most cases the parameter A, which varies most rapidly with temperature, is the deciding factor.

The following ranges may be considered: (i) $z \geqslant 6$; then $z \propto \log t$ and $1/\delta \propto \log t$. Over a short range of values of δ this is equivalent to the logarithmic law of growth, $\Delta\delta \propto \Delta \log t$, although $\Delta(1/\delta) \propto \Delta \log t$ is a better approximation. (ii) $6 > z > 2$; no simple function relates z and t. (iii) $z \sim 2$; a cubic law of growth holds approximately over a narrow range of values of t. (iv) $z \leqslant 1$; a parabolic law of growth $\delta^2 = 4ABVrt$ if $\delta = 0$ at $t = 0$.

Gulbransen and Wysong (1947) studied the oxidation of aluminium in air at temperatures between 350°c and 550°c. At 400°c a film of thickness 265 Å is obtained in 100 minutes, and at 500°c a film of 1220 Å in the same time. Assuming as above an approximate value of 2 v for the potential difference between the two surfaces of the oxide layer, this gives z values of the order of 1–0·2. We should then expect a parabolic law. Gulbransen and Wysong do in fact find that a parabolic law fits their oxidation curves in the temperature range 350–450°c. Their results can be represented by the equation $w^2 = Kt$.

K varies from one specimen to another by a factor of the order of 3. At

450°c, $K \sim 5 \times 10^{-15}$ (g cm^{-2})^2/sec. This experimental formula may be compared with the theoretical parabolic formula, written in the form $w^2 = 4ABVr\rho^2 t$, where ρ is the density, so that $K = 4ABVr\rho^2$. The variation of K with temperature follows the same law as does A, namely $K \propto \exp(-U/\boldsymbol{k}T)$, but the value of U obtained by Gulbransen and Wysong from their experiments at about 400°c is lower than that calculated from A at room temperature, viz. 1 ev as against 1·55 or 1·8 ev. Whether this difference is real, or arises from errors in the experimental data, or in the numerical data required to deduce U from the observed value of A, has not been decided.

To compare K with $4ABVr\rho^2$ the variation of A and B with temperature must be determined. In the case of B this presents no difficulty, since according to the theory $B \propto 1/T$. Thus $B \sim 1{\cdot}5 \times 10^{-6}$ at 450°c. Since the value of U is uncertain it is preferable to calculate its value to obtain agreement between K and $4ABVr\rho^2$. Thus if A equals 10^{-18}, as deduced above, $U \sim 1{\cdot}2$ ev, while if A equals $3{\cdot}6 \times 10^{-23}$, as given by Guntherschulze and Betz, $U \sim 1{\cdot}7$ ev. These values are certainly of the right order.

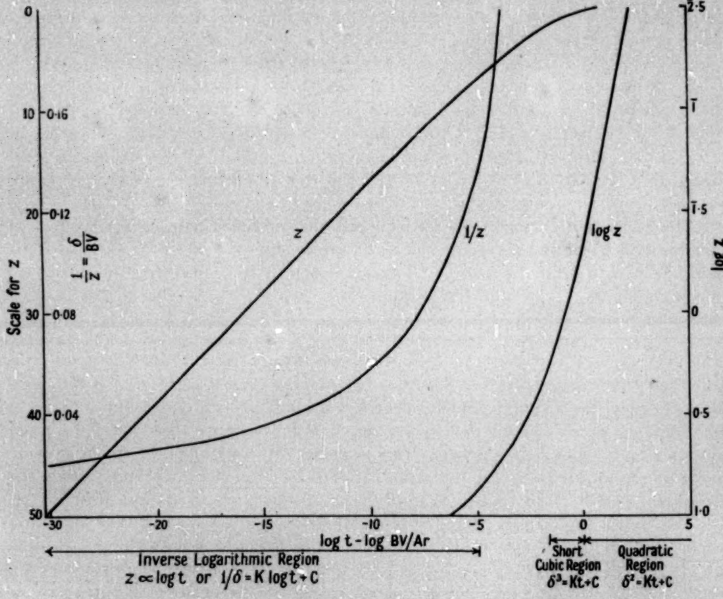


Fig. 9. Theoretical variation of film thickness with time ($\delta = BV/z$).

It follows that the formula relating ion current and electrostatic field in electrolytic oxidation at room temperatures under high potentials can also be used to explain oxidation in air from room temperatures up to about 450°c. Not only does the amended formula give the correct logarithmic formula at low temperatures, and the parabolic at higher, but it predicts correctly the temperature at which the transition takes place, and very approximately the numerical constants involved. The range over which the equation has been tested extends from $z=0{\cdot}1$ to $z\sim 30$, corresponding to a range in time scale of 10^{17}.

ACKNOWLEDGMENTS

The writer wishes to thank Dr. H. M. Finniston and Dr. L. Young of the Metallurgy Division, Atomic Energy Research Establishment, Dr. Van Geel of the Philips Research Laboratories, Eindhoven, and Mr. Gaut of The Plessey Co. Ltd. for helpful discussions. He would like to record his appreciation of experimental work carried out by Messrs. J. J. Polling, A. H. Tilsley and L. M. Neal in connection with this research. He is grateful to the Director, Atomic Energy Research Establishment, Harwell, for permission to publish this paper.

REFERENCES

BETZ, H., 1933, *Z. Phys.*, **82,** 644.
BURGERS, W. G., CLAASSEN, A., and ZERNIKE, J., 1932, *Z. Phys.*, **74,** 593.
CABRERA, N., and MOTT, N. F., 1948–49, *Rep. Prog. Phys.* (London : Physical Society), **12,** 163.
COURSEY, P. R., and RAY, S. N., 1939, *J. Instn. Elect. Engrs.*, **85,** 107.
DEKKER, A. J., and VAN GEEL, W. CH., 1947, *Philips Res. Rep.*, **2,** 313.
DERYAGIN, B. V., and FRIEDLAND, R. M., 1948, *J. Tech. Phys. USSR*, **18,** 1443.
FOWLER, R. H., and NORDHEIM, L., 1928, *Proc. Roy. Soc.* A, **119,** 173.
GRUNERT, H., 1934, *Z. Phys.*, **91,** 49.
GULBRANSEN, E. A., and WYSONG, W. S., 1947, *J. Phys. Coll. Chem.*, **51,** 1087.
GUNTHERSCHULZE, A., 1937, *Elektrolytkondensatoren* (Berlin : Krayn).
GUNTHERSCHULZE, A., and BETZ, H., 1931 a, *Z. Phys.*, **68,** 145; 1931 b, *Ibid.*, **71,** 106; 1932 a, *Ibid.*, **73,** 580; 1932 b, *Ibid.*, **74,** 681; 1932 c, *Ibid.*, **78,** 196; 1934 a, *Ibid.*, **91,** 70; 1934 b, *Ibid.*, **92,** 367; 1935 a, *Ibid.*, **96,** 686; 1935 b, *Ibid.*, **100,** 539; 1937, *Ibid.*, **107,** 347, 633.
GUNTHERSCHULZE, A., and KELLER, F., 1932, *Z. Phys.*, **75,** 78.
MOTT, N. F., 1947, *Trans. Faraday Soc.*, **43,** 429.
VAN GEEL, W. CH., 1931, *Z. Phys.*, **69,** 765.
VAN GEEL, W. CH., and EMMENS, H., 1933, *Z. Phys.*, **87,** 220; 1934, *Physica*, **1,** 415.
VERWEY, E. J. W., 1935 a, *Physica*, **2,** 1059; 1935 b, *J. Chem. Phys.*, **3,** 592; 1935 c, *Z. Kristallogr*, **91,** 65, 317.

RESEARCH NOTES

Radiative Transitions in Germanium

By J. B. Gunn

Elliott Brothers (London) Ltd., Research Laboratories, Borehamwood, Herts.

Communicated by A. E. De Barr; MS. received 19th January 1953

A knowledge of the mechanisms by which an electron in the conduction band of a semiconductor is able to recombine with a hole in the valence band is important in the technology of transistors and similar devices, since it is only with those materials in which recombination proceeds slowly that satisfactory results can be obtained. Two kinds of transition are possible: the non-radiative type, in which the energy E of the hole and electron is dissipated to the lattice vibrations in several stages, and the radiative type, in which recombination occurs in a single stage with the emission of a quantum of radiation of frequency $\nu = E/\boldsymbol{h}$. E should be about equal to the width of the forbidden band, so that ν should correspond to the frequency of the long-wave absorption edge of the lattice. Radiation attributed to the recombination of injected minority carriers in silicon and germanium has been detected by Haynes and Briggs (1952). As no complete account of their work has yet appeared, it is thought worth while to report now the independent discovery of the effect in germanium, even though refinement of the simple technique employed has been prevented by the pressure of other work.

Two similar germanium filaments were used, one as the source of radiation and the other as a photoconductive detector. Each was prepared from n-type material of 6–7 ohm cm resistivity, and, after leads had been soldered to the ends, was etched in warm hydrogen peroxide. Holes were injected into the source filament by an emitter contact placed near one end, and were drawn along its length by an externally applied steady drift field. Since the emitter current was pulsed at 50 c/s, the hole density was also modulated at this frequency. The signal from the detector filament, which was placed with its broad side parallel to, and about 1 mm from, that of the source, was fed to a battery-powered amplifier with a pass band of 30–200 c/s, and thence to an oscilloscope. The output from the oscilloscope amplifier was rectified by a phase-sensitive detector which was synchronized with the emitter current and had an output time constant of 0·5 second. Using an emitter current of 200 mA peak, and a drift current of 10 mA, a signal equal to about twice the noise level was obtained at the detector. The disappearance of this signal on reversing the drift current showed it to be associated with the presence of holes in the source filament, and not the result of thermal radiation from the emitter or other spurious effects. Since it disappeared also when a piece of opaque paper was interposed between source and detector, it must have been due to radiation of a wavelength to which the detector was sensitive, i.e. less than about 1·7 microns.

The sensitivity of the detector was measured by calibrating it against a black body at 650°K and using the curve relating photoconductive response to wavelength given by Goucher (1950). On the assumption that the recombination radiation falls just to the short-wavelength side of the long-wave cut-off of this curve, the flux received by the detector was about 3×10^{11} quanta sec^{-1}. From the geometry of the system, the number of quanta emitted by the source was therefore about $1\cdot3\times10^{12}$ sec^{-1}, which must also have been the number of radiative transitions. Since the emitted wavelength must necessarily be close to that of the long-wave photoconductive cut-off, a precise knowledge of the relative positions of the two spectra is necessary before these figures can be given more significance than that of lower limits, although they are unlikely to be in error by more than an order of magnitude. No correction has been applied for the internal absorption of the source, since the process in which a quantum is emitted and absorbed elsewhere in the crystal represents not a transition but the movement of a hole–electron pair from one place to another. The total number of transitions can be estimated from the change in the conductivity of the source on injection, its dimensions, and the hole lifetime, or from the drift current, the hole lifetime and the hole transit time. In the present case, for which the dimensions were $0\cdot03$ cm $\times$ $0\cdot08$ cm $\times$ $0\cdot85$ cm and the lifetime (measured by the method of Shockley *et al.* (1949)) was 65 μsec, both methods give about 10^{16} sec^{-1} for this quantity. Hence a lower limit is found for the recombination constant for radiative transitions in this specimen of about 2 transitions sec^{-1} per added hole, compared with $1\cdot5\times10^{4}$ sec^{-1} for non-radiative transitions. The agreement of this figure with the theoretical estimate of 1 sec^{-1} given by Shockley *et al.* (1949) must, however, be regarded as fortuitous in view of the errors inherent in the method employed.

I am indebted to Mr. H. Wolfson of Standard Telephone and Cables Ltd., Ilminster, for the germanium used, and to the Directors of Elliott Brothers (London) Ltd. for permission to publish this note.

REFERENCES

GOUCHER, F. S., 1950, *Phys. Rev.*, **78,** 816.
HAYNES, J. R., and BRIGGS, H. B., 1952, *Bull. Amer. Phys. Soc.*, **27,** 2, 14.
SHOCKLEY, W., PEARSON, G. L., and HAYNES, J. R., 1949, *Bell. Syst. Tech. J.*, **28,** 344.

A Note on the Theory of Phase-Contrast Images

BY H. H. HOPKINS
Physics Department, Imperial College, London S.W.7

MS. received 16th December 1952

IN the simple theory of phase-contrast images (Zernike 1942, Françon 1950) the following four simplifying assumptions are made :

(I) the illuminating wave is free from aberration,

(II) the phase-perturbations $\phi(u, v)$ produced by the object are small enough to permit the approximation $e^{i\phi(u,v)}=1+i\phi(u, v)$,

(III) all the direct light, and none of the diffracted light, passes through the phase-changing region of the phase-plate,

(IV) all the diffracted light falls within the aperture of the image-forming optical system.

It is shown in what follows that the restrictive assumptions (I) and (II) are unnecessary, providing (III) and (IV) are made. This extension of the simple theory makes possible the study of images of objects showing large variations of optical thickness.

Let (u, v) be reduced rectangular coordinates in the object plane; that is

$$u = \frac{2\pi}{\lambda}(N \sin \alpha)\xi, \qquad v = \frac{2\pi}{\lambda}(N \sin \alpha)\eta \qquad \ldots\ldots(1)$$

where (ξ, η) are geometrical lengths, N is the refractive index of the object space, and α is the angular radius of the entrance pupil of the optical system. The coordinates (x, y) will refer to the entrance pupil, which is of radius $(x^2+y^2)^{1/2}=1$. These coordinate systems are so defined that $u'=u$, $v'=v$ define the geometrical image of the point (u, v), and the coordinates (x, y) denote both an incident ray and the corresponding ray in the image space. The object and image coordinate systems (u, v), (u', v') have their positive axes in opposite directions, corresponding to the inversion of the image. The positive directions of (u, v) are opposite those of (x, y). The phase-plate is taken to be located at either the entrance or the exit pupil.

If $\Psi(u, v)$ denotes the wave-aberration of the incident wave, and $E(u, v)$ is a complex function which describes the structure of the object, the wave has a complex amplitude

$$G(u, v) = e^{i\Psi(u, v)} E(u, v), \qquad \ldots\ldots(2)$$

after passing through the object. The disturbance at the entrance pupil of the optical system is obtained by finding the inverse Fourier transform of $G(u, v)$. That is, if S denotes the object plane,

$$g(x, y) = \frac{1}{2\pi} \iint_S e^{i\Psi(u, v)} E(u, v)\, e^{-i(ux+vy)}\, du\, dv,$$

which may be written

$$g(x, y) = \frac{1}{2\pi} \iint_S e^{i\Psi(u, v)}\, e^{-i(ux+vy)}\, du\, dv$$
$$+ \frac{1}{2\pi} \iint_S e^{i\Psi(u, v)} \{E(u, v) - 1\}\, e^{-i(ux+vy)}\, du\, dv, \qquad \ldots\ldots(3)$$

the first term of which represents the direct light, the second term representing the light diffracted by the structure of the object. According to assumptions (III) and (IV) above, all of the former passes through the phase-changing area, having complex transmission Ae^{iP}, whereas all of the latter falls outside this but within the region of the pupil. Hence the first integral is zero except for points within the phase-changing area, and the second integral is zero except for points within the remainder of the pupil.

The disturbance in the exit pupil is thus given by

$$g'(x, y) = \frac{A\, e^{iP}}{2\pi} \iint_S e^{i\Psi(u, v)}\, e^{-i(ux+vy)}\, du\, dv$$
$$+ \frac{1}{2\pi} \iint_S e^{i\Psi(u, v)} \{E(u, v) - 1\}\, e^{-i(ux+vy)}\, du\, dv,$$

and the Fourier transform of this gives the complex amplitude at the point (u', v') of the image plane. This is

$$G'(u'\ v') = \frac{A\, e^{iP}}{(2\pi)^2} \iint_\sigma \iint_S \exp\{i\Psi(u, v)\} \exp[i\{(u'-u)x + (v'-v)y\}]\, du\, dv\, dx\, dy$$
$$+ \left(\frac{1}{2\pi}\right)^2 \iint_\Sigma \iint_S \{E(u, v) - 1\} \exp\{i\Psi(u, v)\}$$
$$\times \exp[i\{(u'-u)x + (v'-v)y\}]\, du\, dv\, dx\, dy, \qquad \ldots\ldots(4)$$

where σ denotes the phase changing area, and Σ denotes the area of the pupil excluding σ.

If $G(u, v)$ is understood to be zero outside S, the integrations over (u, v) may be given formally infinite limits. Moreover, in the integrations over (x, y) of (4) the integrands are zero outside σ and Σ respectively. Hence these integrations may also formally be given infinite limits. Equation (4) may then be evaluated by Fourier's theorem to give

$$G'(u', v') = \exp\{i\Psi(u', v')\}\{A\,e^{iP} + E(u', v') - 1\}, \qquad \ldots\ldots(5)$$

within S', the image of S, and zero at points exterior to S'. The intensity in the image is given by the squared modulus of $G'(u', v')$, that is, by

$$I'(u', v') = |A\,e^{iP} + E(u', v') - 1|^2, \qquad \ldots\ldots(6)$$

and is therefore independent of the aberration of the incoming wave.

The above formula (6) applies to the image of any type of object. If this latter is of uniform transparency, and shows only a phase structure, we may write $E(u, v) = e^{i\phi(u, v)}$. Expansion of (6) then gives

$$I'(u', v') = A^2 + 2(A \sin P)\sin\phi(u', v') + 2(1 - A\cos P)\{1 - \cos\phi(u', v')\}, \ldots(7)$$

which is exact, providing assumptions (III) and (IV) are valid. It will be seen that (7) reduces to the result of the simple theory if the approximations $\sin\phi(u', v') = \phi(u', v')$ and $\cos\phi(u', v') = 1$ are made. However, if the phase-plate is made absorbing, as is often the case, A is small and it may no longer be possible to neglect the third term of (7) in comparison with the others, even for moderate values of $\phi(u', v')$.

$A = 0$ corresponds to the case of dark-ground observation, and

$$I'(u', v') = 2\{1 - \cos\phi(u', v')\}, \qquad \ldots\ldots(8)$$

from which it appears that the dark-ground image is independent of the sign of $\phi(u, v)$.

It is possible to use the expression (6) even when the conditions (III) and (IV) are not satisfied. If the complex transmission of the object is Fourier analysed, those frequencies which diffract light within the phase-changing area may be regarded as part of the aberration of the illuminating wave, and are thus not seen in the image. Similarly those frequencies which diffract light outside the area of the pupil do not appear in the image, since that light is lost. Hence (6) will give the intensity distribution in the image of an 'effective' object comprising those frequencies which diffract light within the pupil of the optical system, but outside the phase-changing area of the phase-plate. Of course this 'effective' object may show a variation of transparency, even though the actual object $E(u, v)$ shows only differences of phase. When this occurs, phase structures will appear visible even in the focused images of phase objects and using no phase-plate in the system (Hopkins 1952).

It may be noted that the double Fourier transform may be replaced by the Hankel transform of zero order if the object has radial symmetry, and considerations obtain which are exactly parallel to those given above.

REFERENCES

FRANÇON, M., 1950, *Le contraste de phase* (Paris : Editions de la Revue d'Optique).

HOPKINS, H. H., 1952, *Colloque sur le contraste de phase* (Paris : Editions de la Revue d'Optique).

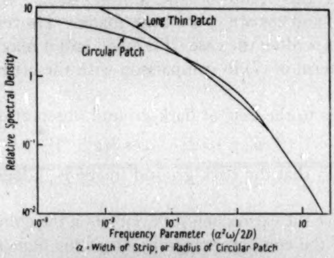
ZERNIKE, F., 1942, *Physica*, **9**, 686.

LETTERS TO THE EDITOR

Contact Noise in Semiconductors

In his analysis of noise in current-carrying semiconductors Macfarlane (1950) considered the fluctuations due to the surface diffusion of ions into and out of patches in which modulation of the current flow by the ions is produced. The patches were assumed to have the form either of a long thin strip or of a circle, and the modulation effect of an ion within the patch was assumed to be independent of its position therein. The problem considered was thus a special case of the much more general class of diffusional noise discussed by Richardson (1950).

It is the purpose of the present letter to point out that the result for the strip can be expressed in a simple closed form and that the analysis for the circle is in error but, when corrected, it can also be expressed in a closed form. When so corrected, however, the spectrum due to a circular patch is found to behave quite differently from the approximate $1/f$ form derived from the erroneous analysis.



Comparison of the forms of the spectral density due to ionic diffusion across a long thin patch and a circular patch.

For a long thin patch the integral of Macfarlane's equation (17) can be readily evaluated as

$$F(p)=\frac{1-\exp\{-(2p)^{1/2}\}\{\cos(2p)^{1/2}+\sin(2p)^{1/2}\}}{(2p)^{3/2}}$$

in Macfarlane's notation.

In the case of the circular patch of radius a the probability P of the escape of an ion from the patch in time τ is given by $P=\mathrm{e}^{-x}[I_0(x)+I_1(x)]$ where $x=a^2/2D\tau$ with D as the surface diffusion constant; Macfarlane's equations (5) and (6) give an incorrect value for the parameter x. The spectral density of the noise arising from a circular patch is then found to be of the form

$$-(\mathrm{ber}_1 z\,\mathrm{kei}_1 z+\mathrm{bei}_1 z\,\mathrm{ker}_1 z)/z^2$$

where $z=(a^2\omega/D)^{1/2}$. At low frequencies the spectral density tends logarithmically to infinity although, of course, the integrated noise over all frequencies is convergent; at high frequencies the noise decreases as $f^{-3/2}$. There is no appreciable range of frequency over which the spectral density can be regarded as approximating to a $1/f$ dependence and in fact, as the figure shows, the circular patch gives a spectrum less resembling the $1/f$ form than that for the long thin patch (a is the width of the strip or the radius of the circular patch).

ACKNOWLEDGMENT

The work described above was carried out as part of the programme of the Radio Research Board, and this note is published by permission of the Department of Scientific and Industrial Research.

Department of Scientific and Industrial Research, **R. E. BURGESS.**
Radio Research Station,
Ditton Park, Slough, Bucks.
12th December 1952.

MACFARLANE, G. G., 1950, *Proc. Phys. Soc.* B, **63**, 807.
RICHARDSON, J. M., 1950, *Bell Syst. Tech. J.*, **29**, 117.

REVIEWS OF BOOKS

Introduction to Concepts and Theories in Physical Science, by GERALD HOLTON. Pp. xviii + 650. (Cambridge, Mass.: Addison Wesley Press, 1952.) $6.50.

The standard university courses in physics as a main or an ancillary subject at British universities have frequently been criticized on the grounds that they lay too much emphasis on imparting a vast amount of factual material to the students instead of concentrating on the thorough teaching of fundamentals. Only at few places does the curriculum include special introductory courses on the fundamental aspects and their historic development, and on scientific method and philosophy of science. The need for such courses has been generally recognized in the U.S.A., and the author, who has great experience in teaching physics and general education in the physical sciences, has written this book as a guide to teachers and students of courses of this type. To quote from his preface: "At the end of such a course the student will be in a position to know the principal laws and the evolution of key conceptual schemes, but it is then also hoped that he, as a responsible citizen, will understand the criteria of validity in scientific thought, the conditions that aid the fruitful growth of science—and perhaps even the exhilaration that binds his instructor to the scientific profession."

The book is most likely to achieve this aim, as it is written with great skill, care and originality. No previous knowledge of the subject and only the most elementary knowledge of mathematics is required. It contains many interesting problems and at the end of each chapter a carefully selected list of books for further reading. The main emphasis is laid on the development of the fundamental ideas of the dynamics of particles including the laws of planetary motion, and the origins of the atomic theory of physics and chemistry, terminating in the concept of the nuclear atom and the fundaments of quantum theory. A whole section is devoted to a discourse on structure and method in physical science.

It is perhaps regrettable that the continuum aspect of physical phenomena has been pushed into the background, so much so that the classical physics of electricity and light is confined to the first two chapters of the section on quantum theory and the nuclear atom. After all, the discontinuum and the continuum aspects have vied with each other throughout the whole history of physics, and the current theories are made up of elements from both these concepts in about equal parts. It is also doubtful whether illustrations of the classical conservation principles drawn from atomic and nuclear physics at a stage where the reader has not yet been acquainted with the basic facts of atomic physics are serving a useful purpose unless it is supposed that the students attending the course know a great deal about electrons, protons and nuclear reactions from popular magazines, but practically nothing about Newton's laws. The very casual introduction of the concept of probability and its role in the kinetic theory is also in sharp contrast to the great thoroughness with which the classical concepts of dynamics are discussed.

Notwithstanding these shortcomings the book can be warmly recommended, and it is to be hoped that introductory courses on similar lines will soon be generally available to students in this country. R. FURTH.

Sir James Jeans—a Biography, by the late E. A. MILNE, with a Memoir by S. C. ROBERTS. Pp. xvi + 176. (Cambridge: University Press, 1952.) 21*s*.

To praise famous men is a pleasant thing, and the biography of James Jeans by the late E. A. Milne is a pleasant book to read. The reviewer in his present capacity praises two great men, for Milne has written a book in spirited, at times sprightly, style, most readable and, as becomes a biography, both revealing and informative.

No more suitable biographer than Professor Milne could have been chosen, for he was himself a great mathematical physicist and, like Jeans, was a reliable contributor to astronomy and astrophysics; moreover, the twain had on more than one occasion crossed swords and fought battles of moment in their time.

Long before Jeans became the interpreter of the Universe to the masses he was, of course, a distinguished scientist, a leader in the first rank, and the manner of his entry—or was it enticement ?—into popular exposition is well outlined by Mr. S. C. Roberts of the Cambridge University Press, in a quite delightful ' Memoir ' which, in the nature of a preface, precedes chapter 1 of the book.

This Memoir is packed with detail which reveals the man rather than the scientist in Jeans, and no one reading these all too few seven pages can doubt that Mr. Roberts was the only man to write them. Jeans was so shy and retiring that few knew him or could pass, or have wished to pass, the barriers which extreme shyness raises. Many will learn for the first time the circumstances of his second romance in later life which led to his second happy marriage to Susi Hoch, a young Viennese musician. A charming story well told.

Professor Milne in twelve well-arranged chapters told the story of Jeans' life: from the more personal angle in the first six chapters, and summarizing the contributions to knowledge in fascinating style in the second six.

Jeans was a precocious child, we are told, and could tell the time at the age of three and could read at four; he got full encouragement from his journalist father. The child factorized cab numbers, learned columns of mysterious logarithms by heart, was exercised prodigiously by perpetual motion and wrote a book (bound prophetically in light blue) entitled *Clocks, by J. Jeans.* It is interesting to learn that his contemporaries at Merchant Taylors' School were afterwards astonished at Jeans' vogue as a popular writer in view of his school reputation that " he could never see that anything needed explaining ".

The chapter on Cambridge will interest all lovers of that great place of learning, reminding the reader of the old days of the Senior Wrangler and the amount of daily work—in Jeans' estimation—required to win the title !

He—with other distinguished names—took the first part of the Tripos in two years (1898), and after a further two years, interrupted by illness, took Part II.

A Smith's Prize came to him in 1901 for an essay entitled ' The Distribution of Molecular Energy '. Thus early did his genius proclaim itself and its trend.

In 1904 his famous book *Dynamical Theory of Gases* appeared—written incidentally during enforced leisure at sanatoria.

1905 saw publication of his definitive solution of the problem of the partition of energy between matter and radiation; his solution was in utter contradiction to experiment, and so led the way to the acceptance of Planck's quantum theory. Jeans' re-derivation of the Rayleigh formula revealed an error in the constant, so that the formula ($8\pi \boldsymbol{R} T\lambda^{-4}d\lambda$) became known as the Rayleigh–Jeans formula.

Few know how it came about that Jeans went to Princeton in 1905. The manner of his going, the works he published while there, and his marriage to a charming American woman have all been duly recounted in this very full biography.

It was during his four years in the States that Jeans, at the early age of 28, was elected a Fellow of the Royal Society, to the pleasure of his friends and the delight of the students, who lightly sang :

Here's to Jimmy Hopwood Jeans,
He tries to make us Math-Machines,
A young and brilliant F.R.S.,
That's going some, we all confess.

Returning to England, Jeans stayed in Cambridge for a few years as Stokes Lecturer but retired in 1912, and at Guildford rapidly made Physical History.

In 1914 appeared his famous Report on *Radiation and the Quantum Theory* (Physical Society) which, after the war, when people had time to read it, did much to establish the theory and help Bohr, in his unorthodoxy, to establish his epoch making views. A number of pages are devoted to a quite fascinating summary of the Eddington–Jeans controversy—into which Milne himself entered—on the role of radiation pressure in star equilibrium. It is well worth reading. An important decade of Jeans' life was his ten years' tenure of the Secretaryship of the Royal Society, 1919–29, a most influential and time consuming post, but one which did not stop Jeans working and taking interest in atomicity and quanta (Rouse Bell Lecture 1925) and in general cosmogony (Adams Prize Essay : Problems of Cosmogony and Stellar Dynamics (1919) and Astronomy and Cosmogony (1928)).

It is curious to reflect that this last work in 1928 was his last, or nearly his last, technical contribution, since its final chapter contained such striking sentences as the following, which arrested instantly Mr. Roberts' alert attention :—" Let us, however, reflect that mankind is at the very beginning of its existence ; on the astronomical time-scale it has lived for only a few brief moments and has only just begun to notice the cosmos outside itself. It is perhaps hardly likely to interpret its surroundings aright in the first few moments its eyes are open."

An exciting ' popular book ' appeared in 1929, the famous *The Universe Around Us,* followed in 1930 by *The Mysterious Universe,* in the nature of a sequel —books which one hopes are still read ! They are products of a great mind which was no longer concerned with technical detail but preferred to stand back and survey what " he himself and others had accomplished and interpret it for the benefit of the intelligent non-specialist ".

Some 50 pages of the biography are devoted, very suitably, to sketches in " as non-technical language as possible of the topics to which Jeans, as a mathematical physicist, had, in his earlier, productive years, devoted himself ".

Suffice it to say that the biographer, with his special knowledge, has made a very good and entrancingly interesting job of it ! It would be a mistake to do more than mention the chapter headings : Rotating fluid masses, Star clusters, The equilibrium of the stars, and the last chapter entitled Jeans and Philosophy.

Professor Milne was never one to accept *ex cathedra* statements quiescently; his was a critical mind constantly in mesh so that it is not surprising that the gently critical vein of some of his earlier pages should mount to something like a crescendo in this last chapter on Jeans' philosophy. There are few physicists and, I imagine no theoretical or mathematical physicists who have never crossed that nebulous and exciting frontier between physics and philosophy. Most return as quickly as may be or say little of their thoughts or feelings. It was otherwise with Jeans, and also, in a different way, Milne. These are some of the things said in the last chapter of the biography : " . . . Jeans' use of philosophy is disappointing " ; " He never brought his critical powers to bear on the theory of relativity "; " he wrote facilely of expanding space when it was the fashion . . . without enquiring what in the world this could mean "; " he wrote of the nebulae as ' straws showing which way the streams of space were following ', a metaphor which, however poetical, is more calculated to darken counsel than to enlighten our minds "; " his innate reverence for mathematics led him to consider the Creator of the Universe as in essence a mathematician ".

There is a most excellent bibliography, including all Jeans' published scientific papers, books and lectures, and a very helpful index.

No one reading this outstandingly well written biography can fail to have his attention gripped and his interests aroused. There is a good deal of criticism interspersed in its 166 pages, but it is nowhere so bitter as to call to the reader's mind *de mortuis nil nisi bonum.* R. WHIDDINGTON.

Symposium on Radiobiology : the Basic Aspects of Radiation Effects on Living Systems, edited by JAMES J. NICKSON. Pp. xii + 465. (New York : John Wiley; London: Chapman and Hall, 1952.) 60*s.*

The biological effects of radiation are becoming increasingly important as a field of study of significance not only to the individual but also to the population as a whole. The present volume is accordingly a welcome review of the state of knowledge of the mechanism of the action of radiation on living systems.

The book comprises the 23 scientific papers presented at a symposium, held at Oberlin College in June 1950, and sponsored primarily by the National Research Council of the U.S. National Academy of Sciences. The objective of the symposium was the presentation of a comprehensive analysis of the fundamental radiobiological concepts concerning the effects of radiation on living cells, by a group of acknowledged authorities, all specialists in their respective fields: the list of contributors included W. M. Dale, G. Hevesy and R. Latarjet from Europe. The papers are essentially reviews of the position in a particular branch of the subject and the presentation of original data has been minimized.

Of the 23 lectures, the first four concern the ionizing effects of different types of radiation including high energy beams and neutrons; these are followed by five lectures on the chemical effects of radiation in gaseous and aqueous systems and on elementary radiobiological chemistry. Consideration is then given to the effect on cell metabolism and biochemical substances in cells, to the physical and chemical factors affecting the sensitivity of cells to radiation and to molecular effects. The next group of lectures deals with chromosome aberrations and the factors affecting radio-sensitivity and is followed by a study of spontaneous and radiation-induced gene mutations and their effects on populations. We then have some penetrating speculations on the cellular action of radiation and a discussion of the target theory and intermediate action theory of the action of radiation on cells, in the light of evidence of the effects of irradiation, and particularly of specific ionization, on yeast cells. The book concludes with a group of four lectures on the effects of irradiation on mammals covering the physiological factors in radiation injury and lethality, the biological effects of quality and intensity of irradiation and the implications of the extensive work on the radiation genetics of mice populations. In most instances an interesting brief discussion of the paper is also included and most of the papers are furnished with extensive bibliographies.

The objectives of the symposium appear to have been very amply attained in this series of lectures. The book has both the advantages and disadvantages of a symposium: the divergency of viewpoints is stimulating, but there is necessarily considerable overlap in the papers. There is also diversity in the depth of treatment, but several of the papers provide excellent and penetrating reviews which will be of considerable interest to those engaged in research on the biological effects of radiation. The book deals effectively with the complexity of this subject which involves so much radiation chemistry, cytology and genetics as well as physics. The physicist will find the papers on the cognate fields absorbing and lucid but he will not find as much as the biologist or radiation chemist to provoke a research interest. It seems unfortunate that a volume which presents the state of knowledge on a rapidly expanding subject should take over two years to publish, but the present volume provides a valuable supplement to the existing standard works on radiobiology. It is perhaps noteworthy that the papers cover the same field as certain of the papers in the report (also recently published) of the Conference on the Biological Hazards of Atomic Energy convened by the Institute of Biology and the Atomic Scientists Association and held in London in October 1950. W. G. MARLEY.

Electronic Analog Computers, by **Gramino A. Korn** and **Theresa M. Korn**. Pp. xv + 378. (New York, London : McGraw-Hill, 1952.) $7.00, 59*s*. 6*d*.

This book will form a useful addition to the literature of analogue computing. It is perhaps unfortunate that the authors have confined themselves almost exclusively to the problem of the design of differential analysers but, even with this limitation, the book contains many ideas which will be useful in other fields of computation.

In the first chapter the fundamental computing elements are considered in a general way. Networks suitable for the operations of addition and subtraction are given and multiplication by means of a servo driven potentiometer is shown to be an applicable technique at low frequencies. The differentiation and integration of a function is next taken up and it is indicated that differentiating circuits have not been much used in serious analogue computation.

The second chapter discusses the set-up procedure favoured by the authors and is mostly devoted to the questions of scale factors and the replacement of the differential operation by integrations. Several equations of interest to physicists are discussed in detail and solutions obtained on a computer are shown for the equations of Mathieu, Legendre and van der Pol.

For the physicist more interested in applications than in computer design, the third chapter will be most stimulating. Practical problems discussed include dynamic vibration dampers, the simple servo, the spherical cannon ball with air resistance and the equations of flight. Flight simulators receive considerable attention and it is shown how a suitable computer can be used in the design and testing of automatic pilots. The chapter concludes with an all too brief discussion of simultaneous equations, integral transforms and boundary and eigenvalue problems.

Chapter four sees the start of the really practical part of the book. The setting of constants and the multiplication of functions by constants are discussed, chiefly with reference to simple potentiometers. This is followed by a detailed analysis of the R.C. integrator and of the improvement of its characteristics by feedback. The chapter concludes with a discussion of the stabilization of amplifiers by feed back.

At the start of the fifth chapter the statement appears (p. 167) " . . . . an exhaustive treatment of amplifier design is entirely beyond the scope of this book ". This is surprising since the remainder of the chapter contains an exhaustive account of the design of d.c. amplifiers. The reviewer found the account of drift compensation particularly valuable especially as no other comprehensive treatment appears to be readily available. An omission, which might well be remedied in any future edition, is that of all mention of magnetic amplifiers.

Chapter six discusses techniques for multiplication and for function generation. There does not appear to be anything of particular novelty in any of the circuits given but it is satisfactory to have the various techniques gathered together in one place.

Auxiliary circuits for setting initial conditions and available output mechanisms form the chief subjects of chapter seven, and a brief description of power supply design and of checking procedures ends the chapter.

The final chapter describes some complete computer installations which are available in the United States of America. A sound discussion of design considerations for future machines is followed by a somewhat sketchy discussion of accuracy. Those concerned with the actual building of computers, especially in this country, will sympathize with the statement that the computer room should be a 'showplace'.

An appendix gives a mathematical discussion of the properties of parallel feedback type operational amplifiers. There is a reasonably complete bibliography and good author and subject indexes.

The reviewer noticed a few errors, the penultimate line of p. 12 should have R_0 instead of R_1. The footnote on p. 130 appears to refer to a further appendix which has not, in fact, been included, and, in the reviewer's experience, it seems most unlikely that electro-mechanical relays are available for operation at 1500 c/s as stated on p. 228. A. D. BOOTH.

Metallurgical Equilibrium Diagrams, by W. HUME-ROTHERY, J. W. CHRISTIAN and W. B. PEARSON. Physics in Industry Series. Pp. 311. (London: Institute of Physics, 1952.) 50*s*.

The equilibrium diagrams of alloy systems occupy a very important position in metallurgical knowledge and an increasing need has been felt in recent years for a comprehensive account of the methods available for their determination. The present book meets this need and will be especially welcome as coming from the laboratory in which Dr. Hume-Rothery has built up a tradition of work of the highest quality in this field. Although written primarily for metallurgists it contains much that could be read with profit by anyone—physicist, chemist, or engineer—who is concerned with measurements of the properties of alloys. The days are fortunately past when physicists would report the values of some property across an alloy system and discuss its variation without reference to changes of constitution, but many papers still reveal a lack of appreciation of the precautions required in the preparation and heat treatment of alloys if measurements on them are to have real significance.

The book opens with a survey of the general principles and thermodynamic basis of phase diagrams, which includes a brief consideration of the status to be assigned to order–disorder changes, without, however, showing the forms that can be taken by superlattice phase fields. The discussion of general experimental procedure that follows includes a useful chapter on refractory materials, and a table showing the appropriate crucible materials and atmospheres for the melting of different metals, which should prove invaluable to a wide range of workers. The chapters on thermal analysis are remarkable in being the only description of these methods known to the reviewer in which there are no errors in the diagrams illustrating the interpretation of heating and cooling curves. The application of x-ray methods to the determination of diagrams is discussed critically, and attention is drawn to possible sources of error; the authors recognize the usefulness of such methods, properly applied, and it is to be hoped that one has heard the last echoes of the 'x-ray versus microscope' controversy. Other physical methods that can be used in this type of work are discussed only briefly. The section on ternary systems that follows is an excellent exposition of this involved subject and draws much of its strength from the remarkably well produced diagrams which make clear many of the points which are difficult to visualize.

The book is described as being intended for fourth-year honours students and research students in metallurgy, but much of the subject matter is appropriate to the earlier stages of metallurgy degree courses. In addition the sections on the preparation and heat treatment of alloys and the emphasis laid on the need for chemical analysis of actual specimens should prove useful to any scientist concerned with the properties of alloys.

It should be made clear (for the title fails to do so) that the observed phase diagrams of actual alloy systems do not lie within the scope of this book.

B. R. COLES.

Charbons activés, by C. COURTY. Pp. ix + 534. (Paris: Gauthier-Villars, 1952.) 4500 fr.

This book is a treatise of 514 pages on the adsorptive properties of activated charcoal. An admirable balance is attained between the description of experimental work, a thorough account of the underlying physics and chemistry, and a mathematical treatment of the theories of adsorption. The author admits that neither the investigations nor the theories are by any means complete, but he makes a valuable contribution towards a full understanding by offering a very careful and critical survey of what has so far been achieved. One of the main problems is the mechanism of the chemical sorption of gases and vapours, and here M. Courty's own researches have played an important part in the progress which has been made in recent years. It is evident throughout that he is writing with authority.

The book opens with an account of surface tension and capillarity, and a critical discussion of the various theories of adsorption. A similarly careful survey follows, dealing with the types of heat of adsorption and the kinetics of adsorption. While these treatments are quite general, special attention is naturally given to those theories which can be applied to highly activated porous surfaces, and to charcoal in particular. The remainder of the book is more concerned with experiment. Measurements of the sorption of various gases and gas mixtures are described, followed by the properties of charcoal relevant to its use in respirators. These include studies of the desorption processes, and the correlation between sorption and numerous physical properties, such as grain size, density and porosity. The book concludes with a detailed account of work on chemical sorption, and on methods of testing the efficiency of sorption and the degree of activation of the charcoal.

The standard of production is reasonable as regards legibility of type, mathematics and diagrams. The book is, however, paper covered, which does not offer a high degree of permanence. Moreover it is very irritating to the purchaser to be supplied with a book in which almost all the pages are uncut. The writer estimates that he spent almost two hours in page-cutting. This standard of production does not seem consistent with the price of the book, and compares unfavourably with standards in this country. D. A. WRIGHT.

Astrophysics. A Topical Symposium, edited by J. A. HYNEK. Pp. xii + 703. (New York, London: McGraw-Hill, 1951.) 102*s*.

This composite work by fifteen authors is issued to commemorate the fiftieth anniversary of the Yerkes Observatory. It forms a fairly comprehensive review of astrophysics with some reference to its growth during the last fifty years.

Each author has been given freedom to cover his subject in his own particular style and hence one cannot expect the various chapters to fit into one another in one continuous theme. There is some overlapping in the sections on normal stellar spectra and on the physics of stars. Certain wide sections of astrophysics are untouched, but in the fields that are covered it can be used somewhat as a textbook. In this respect it is sure to command wide popularity because no general astrophysical textbook at the research level exists in English at present.

The book is appropriately introduced by B. Stromgren with an account of the development of astrophysics during the last half century. It is then divided into four main parts containing a total of fourteen chapters by individual authors.

Part one, on spectroscopic astrophysics, contains matters relating to stellar spectra and stellar atmospheres. P. C. Keenan and W. W. Morgan deal with the significance of stellar classification and discuss the related physical conditions of stars but not descriptive details of their spectra. Questions of stellar line intensities, line structure, curve of growth, and chemical composition are clearly set out by L. H. Aller, while the more individual stellar problems relating to emission lines, rotation, abnormal intensities, gaseous rings and stellar formations (but not flare stars) are made interesting and reasonably systematic by O. Struve. The subject of molecular spectra is difficult to treat in a confined space and P. Swings' article is necessarily devoted to particular bands and compounds present in various cosmic sources rather than to general principles. B. Stromgren has contributed a stimulating survey of the problems of radiation in stellar atmospheres and stellar interior. His article give the full historical development of the subject and is well suited to a commemorative volume although differing in style from the rest of the book.

Part two, on the solar system, reflects closely the interests of the authors and is not in any sense complete. E. Pettit's chapter on the sun and stellar radiation contains the radiation and dimensional data of the sun and a summary account of solar activity. N. T. Bobrovnikoff deals with modern physical problems of comets, particularly the forces that come into play to produce the coma and tails. The problem of their origin is carried further into G. P. Kuiper's chapter on the Origin of the Solar System, which is however too advanced and speculative to suit the rest of the book. It deals with the break up of the solar nebula by turbulence and the formation of planets and satellites.

The third part deals systematically and rather uniformly with binary and variable stars—visual binaries by G. van Biesbroeck, spectroscopic binaries by J. A. Hynek, eclipsing binaries by N. L. Pierce, and intrinsic variables by C. Payne-Gaposchkin. Parallax problems are added to van Biesbroeck's article.

The fourth section has been called the physics of cosmic matter. The rapidly developing subject of interstellar matter is well handled by J. L. Greestein who discusses the physical condition of the gas and cloud formations and also their effect on the stars seen through them. S. Chandrasekhar closes the book with a section on stellar structure in his usual clear cut style.

It will be noticed that there is no attempt to make a complete coverage of astrophysics. There is no treatment of galactic structure, stellar motions, external galaxies, clusters, or planetary nebulae. Other gaps could be found. However the articles reach a high standard and give an up to date view of the major parts of astrophysics. The book is well bound and printed. Although somewhat expensive it is already being well read and it should achieve a wide distribution.

C. W. ALLEN.

CONTENTS OF SECTION A

PAGE

Prof. R. E. Peierls. The Atomic Nucleus and Its Constituents. (6th Rutherford Lecture) 313

Prof. H. S. Green. Boltzmann's Equation in Quantum Mechanics 325

Dr. J. Friedel. Anomaly of the $M_{IV,V}$ Absorption in Heavy Elements . . 333

Mr. G. Goldring. Internal Pairs in Anisotropic Emission 341

Dr. E. P. George, Dr. J. W. MacAnuff and Dr. J. W. Sturgess. Observations of Extensive Cosmic-Ray Showers Below Ground 346

Dr. S. Levine. The Free Energy of the Double Layer of a Colloidal Particle and the Charging Process 357

Dr. S. Levine. Interaction of Two Parallel Colloidal Plates using a Modified Poisson–Boltzmann Equation 365

Mr. W. E. Williams. Solution of a Modified Poisson–Boltzmann Equation for a Single Plane Double Layer 372

Dr. S. J. Wyard. Absorption Coefficients of Gamma-Rays with Energies between 0·3 and 1·5 Mev 382

Mr. N. R. Steenberg. The Polarization of γ-Radiation from Aligned Nuclei . 391

Mr. N. R. Steenberg. Population Distribution of Nuclei Aligned at Low Temperatures 399

Mr. A. Kelly. Neumann Bands in Pure Iron 403

Research Notes :

Prof. H. S. W. Massey and Dr. B. L. Moiseiwitsch. Calculation of the 1s–2s Electron Excitation Cross Section of Hydrogen by a Variational Method 406

Letters to the Editor :

Prof. J. Reekie, Dr. T. S. Hutchison and Mr. C. F. A. Beaumont. X-Ray Scattering from Liquid Helium 409

Dr. B. Bleaney, Mr. K. D. Bowers and Mr. R. S. Trenam. The Nuclear Electric Quadrupole Moments of Copper 63, 65 410

Dr. D. J. E. Ingram. Paramagnetic Resonance in some Manganese Salts . 412

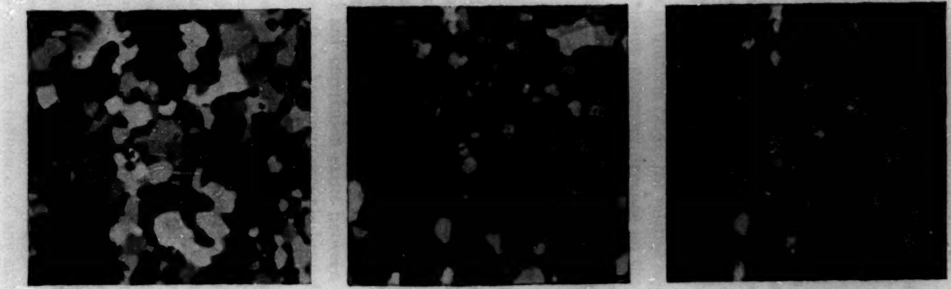
Mr. R. S. Trenam. The Nuclear Magnetic Moment of ^{57}Fe 414

Dr. E. R. Andrew and Mr. R. G. Eades. Separation of the Intramolecular and Intermolecular Contributions to the Second Moment of the Nuclear Magnetic Resonance Spectrum 415

Reviews of Books 417

Corrigendum 420

Contents of Section B 420



(*a*) (*b*) (*c*)

Fig. 1. Effect of rotation of the objective on the grain contrast on uranium.

The Electrical Breakdown of Gases in Non-Uniform Fields at Low Pressure

By F. LLEWELLYN JONES and G. C. WILLIAMS
Department of Physics, University College of Swansea

MS. received 9th January 1953

Abstract. The electrical breakdown of gases between coaxial cylinders at low pressures was studied in air (a gas mixture), hydrogen (a diatomic gas) and in helium (a monatomic gas). The significant secondary ionization processes occurring were correlated with those which are known to occur in uniform fields. The similarity theorem was used to help in the identification of the fundamental collisional processes. The experimental results were consistent with the view that breakdown involves the same primary and secondary ionization processes as in uniform fields: the relative importance of the dominant secondary electron emission processes depends on the nature of the electrodes and the geometry of the system. The effects of traces of impurities in helium were also studied, and the results were accounted for by the action of metastable atoms.

---

§ 1. Breakdown in Non-Uniform Fields

THE electrical discharge between coaxial cylinders at low pressures has been studied for many years, not only because of its practical application in Geiger counters and high-tension power lines, but also because important information about collisional properties of electrons, ions and metastable atoms may be obtained in this way. It is important, also, to investigate the mechanism of the electrical breakdown of gases in non-uniform fields in order to find whether any correlation exists between the significant secondary ionization processes and those known to occur in breakdown under uniform fields (Llewellyn Jones and Davies 1951 a, b.)

In non-uniform fields when the electrode system consists of a wire of radius a and a coaxial cylinder of radius b the field strength at the cathode depends on whether the wire or the cylinder is the cathode. The ratio of the field X_a at the wire to the gas pressure p can be very high ($\sim$300 v/cm/mm Hg) in typical cases. while in regions remote from the wire and near the cylinder the field $X_b (= aX_a/b)$ can be low ($\sim$20 v/cm/mm Hg). The electron energies, and possibly those of ions also, are therefore very much greater in the region near the wire than near the cylinder provided the mean free path is small compared with the radius of the cylinder; the occurrence of charge exchange (Keene 1949, Hasted 1951) and possibly other processes at low values of X/p (Tyndall 1938, Llewellyn Jones 1935) tends, however, to prevent the attainment of high ion energies. Hence, except when such processes occur, the influence of any cathode ionization processes may be compared at different values of X/p, i.e. at X_a/p and X_b/p, simply by investigating breakdown for different polarities of the wire.

Accepted arguments proposed to account for the mechanism of corona breakdown at the lower gas pressures usually run as follows. When the wire is made the cathode and the cylinder the anode all the positive ions move to the wire through an intense field X_a and may thus acquire high energies at the cathode; on the other hand, when the cylinder is the cathode, the positive ions move in the much weaker field X_b near the cylinder, and consequently have low (approximately thermal) energies on reaching the cathode. In general, the coefficient γ of secondary emission due to the impact of positive ions with the cathode increases with X/p; for nickel electrodes in hydrogen, for example, γ increases by a factor of about 10 when X/p increases from 100 to 300 v/cm/mm Hg (Llewellyn Jones 1939). Consequently the value γ_h of the secondary emission coefficient for the wire where the field is high would be at least equal to, and would probably be greater than, the value γ_l for the cylinder, where the field is lower by a factor of about 10. When the secondary ionization process in the breakdown mechanism is due only to the γ process, the Townsend criterion for breakdown in a non-uniform field is

$$1-\gamma\left\{\exp\int_a^b \alpha\,dr-1\right\}=0, \qquad \ldots\ldots(1)$$

where α is the primary coefficient of ionization by single electron collision. The field X_r at any radial distance r at breakdown is related to the breakdown potential V_s by the expression

$$X_r=V_s/r\ln(b/a), \qquad \ldots\ldots(2)$$

and when α/p and γ are both functions of X/p it follows that

$$\gamma=\exp-\left\{aX_s\int_{X_b/p}^{X_a/p}\frac{\alpha/p}{(X/p)^2}\,d\left(\frac{X}{p}\right)\right\}. \qquad \ldots\ldots(3)$$

This gives the dependence of the breakdown potential V_s on the coefficient γ; the higher the value of γ the lower is V_s, and changes in γ produce corresponding changes in V_s. It follows that on this view the breakdown potential V_s^- when the wire is cathode (negative discharge) can be equal to or lower than the value X_s^+ when the wire is anode (positive discharge), but cannot be greater than V_s^+ at the same value of pa. Experiments in the diatomic gases nitrogen (Huxley 1930) and hydrogen (Bruce 1930) show that V_s^- is in fact less than V_s^+. In air, however, the (V_s^-, pa) and (V_s^+, pa) curves intersect (Boulind 1934), V_s^- exceeding V_s^+ when $pa>0{\cdot}7$ mm Hg cm, but V_s^- being less than V_s^+ when $pa<0{\cdot}7$ mm Hg cm. It follows from the above argument, therefore, that the γ process due to impact of positive ions alone cannot account for this result, and that the action of other secondary ionization processes must be considered. It has been suggested that ionization of gas atoms by positive ions (β process) in the space near the wire where the field is high was such a process, and arguments in support of this have been given by Huxley and Bruce (1937) and Townsend (1947); the intersection of the (V_s^-, pa) and (V_s^+, pa) curves was then considered to be a consequence of the relative magnitudes of the β and γ coefficients. The same effect can, however, be produced by other secondary processes, such as the photoelectric effect at the cathode due to the incidence of photons produced in the electron avalanche; the problem of explaining the relative magnitudes of V_s^- and V_s^+ in any given case therefore depends on the identification of the particular secondary ionization processes predominant in the breakdown mechanism.

In pure monatomic gases, on the other hand, no intersection of the (V_s, pa) curves for a positive and a negative discharge has been observed (Penning 1931),

V_s^- being always lower than V_s^+ for helium, neon and argon. When, however, traces of impurities are present the (V_s, pa) curves for positive and negative discharges can intersect (Huxley 1928), a result attributed (Penning 1931) to collisions of the second kind between metastable atoms and the atoms of impurity resulting in the ionization of the atoms of impurity. In the case of positive and negative breakdown in air and oxygen, however, no satisfactory explanation of the intersection of the (V_s, pa) curves has been given, although many attempts have been made in recent years (e.g. Miller and Loeb 1951) to elucidate the mechanism. Investigations by Louw and Naude (1949) and Wiedenbeck and Crane (1949) on spurious counts in Geiger counters, and by Pidd and Madansky (1949) on the large variations in the dead times with changes of the cathode surface in spark counters, indicate that the precise role played by oxide films is not generally appreciated or fully understood. In Geiger counters, for example, the presence of a surface film is common because of the nature of the technique employed in its construction and the nature of the gases which are generally used.

Recent work has shown that thin electro-positive and electro-negative films on the cathode have a profound influence on the breakdown potentials of gases in uniform fields (Llewellyn Jones and Davies 1951 b). Large changes (by factors of the order of 50 times) in the generalized secondary coefficient ω/α representing the total electron production from all causes per ionizing collision in the gas, can be produced by changes in the cathode work function caused by the existence of electro-positive or electro-negative surface layers on the cathode. This is due to the fact that the mechanism of secondary ionization in the discharge is in general not a single process, but generally consists of at least two: cathode emission due to the incidence of positive ions and photoelectric emission due to photons generated in the electron avalanche. The relative importance of these two processes depends on X/p and the geometry of the system. In general, when a number of different secondary ionization processes occur, ω/α can be expressed approximately by the relation

$$\omega = \alpha\gamma + \beta + \delta + \eta + \epsilon, \qquad \ldots\ldots(4)$$

where secondary electron production at the cathode may be produced by the impact of positive ions (γ effect) or of photons (δ effect) and metastable atoms (ϵ effect) and in the gas by ionization by collision by positive ions (β effect) and by photon absorption (η effect) (Dutton, Haydon and Llewellyn Jones 1953). Hence it is necessary to investigate breakdown in non-uniform fields in various gases and to pay particular attention to the influence of surface films on the mechanism, as has already been done for the case of breakdown in uniform fields.

The following sections of the paper describe the experimental investigations of static corona breakdown at low pressures in air (a gas mixture), hydrogen (a pure diatomic gas) and in helium (a monatomic gas) with the object of identifying the secondary ionization mechanisms operating, and then to compare these with the processes known to occur in uniform fields. The method of investigation was to evaluate ω/α as a function of X/p for each case of breakdown, because the shapes of the (ω/α, X/p) curves help to identify the secondary processes; further, application of the similarity and corona relationships gives important information about the nature of the ionization and de-ionization processes. With the three different kinds of gases used, the effects of impurities and of different processes of ionization by collisions of the first and second kind could be examined.

§ 2. APPARATUS AND EXPERIMENTAL PROCEDURE

Three discharge tubes A, B and C were employed, consisting of nickel rod electrodes of diameters ($2a$) 0·159, 0·318 and 0·160 cm respectively, each mounted coaxially in nickel cylinders of internal diameter, ($2b$) 3·945, 3·950 and 2·080 cm respectively, and enclosed in hard glass envelopes all sealed to a common gas supply. The gas pressure was measured with an 'Apiezon' oil manometer, the oil being freed from dissolved gases whenever the nature of the gas was changed. Mercury was excluded from the whole system. The air used was admitted from the laboratory into a reservoir through a calcium chloride drying tube and filtered with glass wool, and was stored over phosphorus pentoxide for at least 24 hours before use. The hydrogen was generated by the electrolysis of a 5% solution of barium hydroxide, dried over phosphorus pentoxide, then diffused through a heated palladium thimble. Commercially produced spectrally pure helium was stored over activated charcoal cooled with liquid oxygen for a period of at least 15 hours before passing into the discharge tubes. The three discharge tubes were always open to a nearby large charcoal trap which was cooled by liquid oxygen when helium was in the tubes. The applied potentials were obtained from a stabilized power pack and measured by a calibrated voltmeter. The discharge current passing after breakdown was recorded by a sensitive microammeter and was limited by a series resistance $\sim 2\mathrm{M}\Omega$. The discharge tubes were outgassed at 300°c–400°c in an electric furnace for some days while being continuously evacuated. For the measurements in hydrogen and in helium the electrode system was subjected to severe hydrogen ion bombardment in order to remove any tenacious surface tarnish films (Llewellyn Jones and Davies 1951 b).

The determination of the breakdown potentials of a gas at different pressures involved difficulties, and many previous observers have not stated clearly what criterion was used to indicate the potential V_s at which breakdown actually occurred. The difficulties are more pronounced for a positive discharge at the higher values of pa (or at the lowest values of X/p), because in this case the ionization, which occurs before breakdown is complete, is confined to the region about the wire, and the loss of electrons to the wire is large. When very small values of the initial current I_0 ($\sim 10^{-14}$ A) are used space charge is negligible, and the self-maintained discharge current I can be increased by a large factor ($\sim 10^8$) without sensibly changing the applied potential. This potential is taken as the breakdown potential V_s.

§ 3. EXPERIMENTAL RESULTS

(i) *Air*

The experimental results for air are best expressed graphically, and the (V_s^+, pa) and (V_s^-, pa) curves obtained with tube C are given in fig. 1. It can be seen that for values of pa greater than 0·7 mm Hg cm, $V_s^- > V_s^+$, but when $pa < 0{\cdot}7$ mm Hg cm then $V_s^- < V_s^+$; these results were reproducible and in general agreement with those obtained by Boulind (1934) for tubes of different dimensions.

Values of ω/α calculated from these values of V_s are given in fig. 2. The calculation of ω/α involves the solution of eqn. (1), in which the coefficient γ is replaced by the more general coefficient ω/α. In order to cover the complete range from X_b/p to X_a/p, values of α/p determined by different observers

(Townsend 1915, Sanders 1933) were used and the integral evaluated graphically. For a negative discharge ω/α increased linearly with X/p, but for the positive discharge ω/α exhibited a peak value at $X/p = 26$ v cm/mm Hg; for values of X/p greater than 40 v/cm/mm Hg ω/α again increased steadily. At a value of X/p of 957 v/cm/mm Hg ω/α was found to be 0·008, which compares favourably with the value 0·01 obtained by Llewellyn Jones and Davies (1951 a) for uniform fields and a nickel cathode.

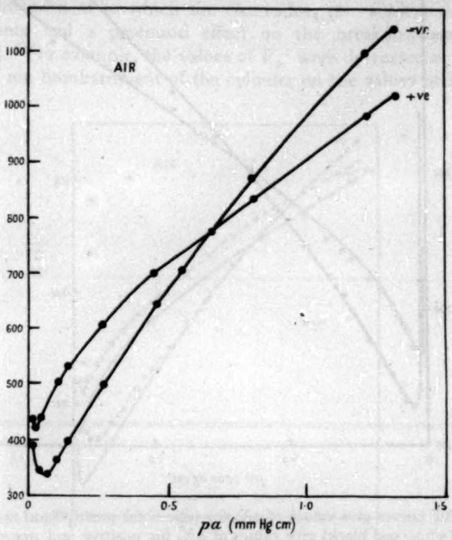


Fig. 1. Breakdown potentials of air for tube C.

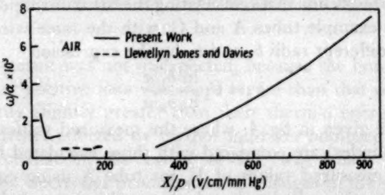


Fig. 2. (ω/α, X/p) curves for air for cathodes of nickel.

Data on the breakdown mechanism may also be obtained from a consideration of the applicability of the similarity and corona theorems. When the similarity theorem

$$V_s = aX_s \ln(b/a) = F(pa) \qquad \ldots\ldots(5)$$

is found to apply to geometrically similar systems, only collisional processes which are functions of X/p play any significant part in breakdown: it is therefore of interest to study under what conditions the theorem applies in air. A set of (aX_s, pa) curves is given in fig. 3, where the continuous curves refer to tube C and the values indicated by white and black circles to tube B. These values lie on the continuous curves throughout the range $0{\cdot}01$ mm Hg cm $< pa < 1{\cdot}6$ mm Hg cm, and this shows that the similarity theorem is satisfied over the whole range of X/p.

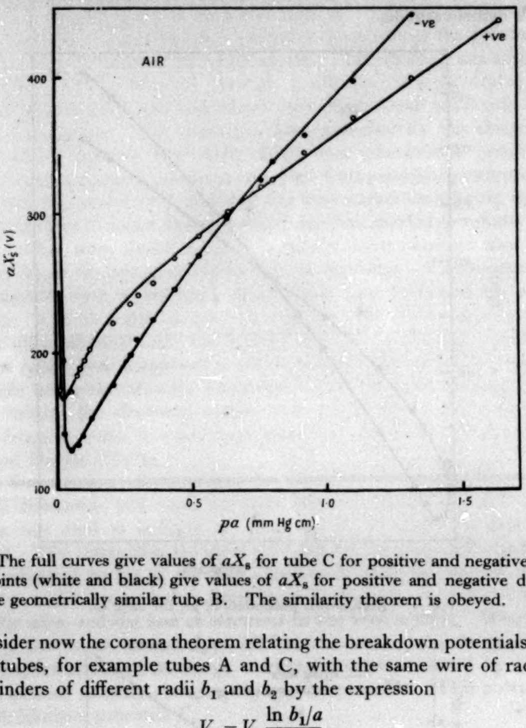


Fig. 3. The full curves give values of aX_s for tube C for positive and negative discharges. Points (white and black) give values of aX_s for positive and negative discharges in the geometrically similar tube B. The similarity theorem is obeyed.

Consider now the corona theorem relating the breakdown potentials V_1 and V_2 for two tubes, for example tubes A and C, with the same wire of radius a, but with cylinders of different radii b_1 and b_2 by the expression

$$V_1 = V_2 \frac{\ln b_1/a}{\ln b_2/a}. \qquad \ldots\ldots(6)$$

Typical results are given in fig. 4, where the measured values of V_s for tube C (white and black circles) are compared with those calculated for C (continuous curves) from the measured values of V_s for tube A using equation (6). The agreement between the calculated and observed values of V_s^- is good when $p < 0{\cdot}4$ mm Hg, but for lower values of p a marked divergence is observed. This was to be expected, since the discharge extended to the cylinder at higher values of pressure in tube C than in tube A, thereby influencing the discharge in C at much higher values of p. For the positive discharge the values of V_s^+, which were measured for tube C, were higher than the calculated values, and this may be explained by the greater part played by diffusion in tube C than in tube A, throughout the range of values of p considered.

(ii) *Hydrogen*

(V_s, pa) data for tube C for different conditions of the electrodes are given in fig. 5. The initial experiments were made with electrodes which had been exposed to air for a period of about 3 months, and the surface of the nickel was therefore coated with a thick ($\gtrsim 10^{-5}$ cm) layer of oxide. At a value of *pa* of 2·23 mm Hg cm the difference between the values of V_s^- and V_s^+ was only 12 v (curves 1 and 2), but this difference increased as *pa* decreased.

The ion bombardment to which the electrodes were subjected during the initial experiments had a profound effect on the breakdown potential V_s (curves 3 and 4). For example, the values of V_s^- were decreased by about 40 v, but the effect of ion bombardment of the cylinder on the values of V_s^+ was not appreciable. This result was not unexpected, because the bombardment of the wire by high energy positive ions was more severe than that of the cylinder by ions of energies only slightly greater than their thermal energies. The values of V_s^- were still further affected by further mild ion bombardment of the wire cathode, but the change in the values of V_s^+ was small. Thus mild ion bombardment of the electrodes produced small changes in the values of V_s^+ but large decreases in V_s^-.

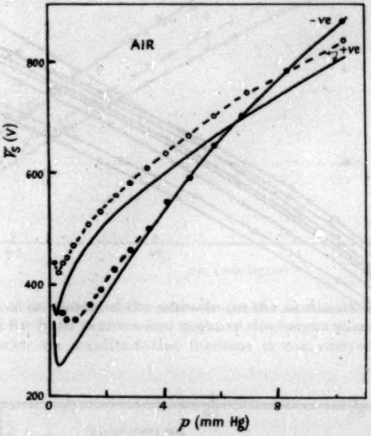


Fig. 4. The measured values (white and black circles) of V_S for tube C are compared with those (curves) calculated from the measured values for tube A using the corona relationship.

At this stage in the experiments air was introduced into the tubes for about one hour during which the electrodes acquired a thin oxide layer. After this the values of V_s^- (curve 5) were found to have decreased by 75 v compared with those of curve 3, but the values of V_s^+ were not affected to any marked degree as a result of the oxidation (curve 6). The tubes were then heated continuously at 300°c while the electrodes were subjected to continuous ion bombardment, and after 10 hours of this treatment they were evacuated for another period of

10 hours while being maintained at 300°c to remove adsorbed gas; pure hydrogen was then admitted by diffusion through palladium and breakdown potential measurements made. The observed values of V_s for three separate sets of measurements with different specimens of gas and with the electrodes in this de-oxidized state all lie, within experimental error, on a single curve (fig. 5, curves 7 and 8).

When the electrodes were de-oxidized, curves 1 and 2 of fig. 6 show that the values of (aX_s, pa) for tube C (continuous curves) coincided with the values

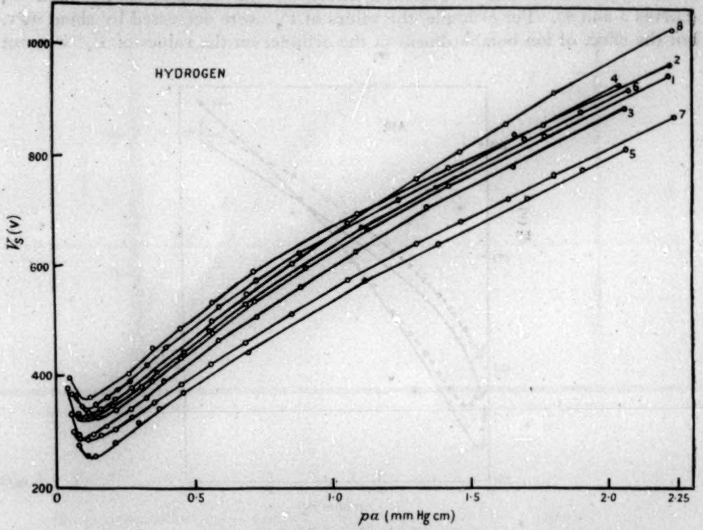


Fig. 5. Breakdown potentials of hydrogen for tube C for different states of oxidation of the electrodes.

1, 3, 5 and 7 negative wire; 2, 4, 6 and 8 positive wire.

(white and black circles) for the geometrically similar tube B, showing that the similarity theorem was satisfied for both positive and negative discharges. When, on the other hand, the electrodes were oxidized the values of (aX_s^+, pa) for tube C (curve 3) and those for tube B (crosses) did not coalesce for values of *pa* greater than 0·5 mm Hg cm, but for values of *pa* less than 0·5 mm Hg cm the theorem was satisfied. Values of (aX_s^-, pa) for tube C (curve 4) and tube B (crosses) satisfied the similarity theorem over the entire range investigated.

Further, it is seen from fig. 7 that the values of V_s measured for tube C (white and black circles) were in agreement with those calculated from the measured values for tube A (continuous curves) over a large range of pressure. At low values of p (<3 mm Hg) the discharge extended to the cylinder in C while the discharge was still remote from the cylinder in A. Hence the measured breakdown potentials for C would be expected to be greater than those calculated from A since the loss of electrons and ions to the cylinder of C was increased. The point at which the values diverged appears at a higher value of p for air than for hydrogen because of the high electron drift velocity in air (Healy and Reed 1941).

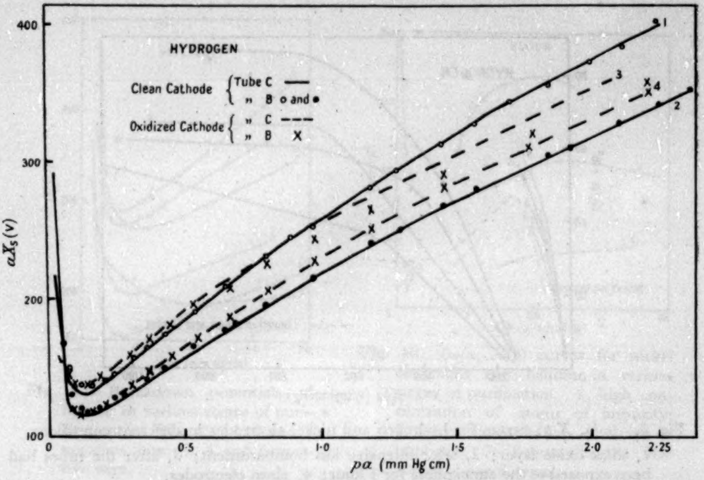


Fig. 6. The influence of oxidation of the cathode on the similarity theorem. The similarity theorem is obeyed for both positive and negative discharges when the electrodes are clean. When the electrodes are oxidized the theorem is not obeyed exactly for a positive discharge.

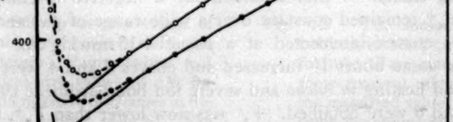


Fig. 7. The measured values (white and black circles) of V_s for tube C are compared with those (curves) calculated from the measured values for tube A using the corona relationship.

For any given value of X/p, ω/α was calculated by solving eqn. (3) graphically, using the values of α/p measured by Hale (1939) and the values of V_s given in fig. 5; these values of ω/α are plotted as functions of X/p in fig. 8 for both positive and negative discharges.

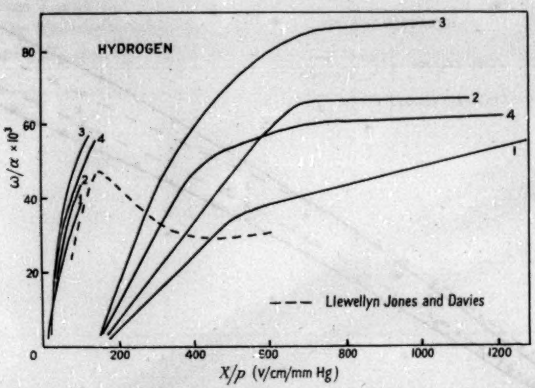


Fig. 8. (ω/α, X/p) curves for hydrogen and nickel electrodes in different conditions. 1, thick oxide layer; 2, after intensive ion bombardment; 3, after the tubes had been exposed to the atmosphere for 1 hour; 4, clean electrodes.

(iii) *Helium*

On completing the investigations with de-oxidized electrodes in hydrogen, the discharge system was evacuated for a period of 10 hours and pure helium was admitted to the desired pressure. After the preliminary outgassing the system still contained small traces of hydrogen adsorbed on the electrodes and the glass envelopes, and typical (V_s, pa) curves for helium in various stages of purification with de-oxidized electrodes are given in curves 1 and 2 of fig. 9 for tube C. The (V_s^+, pa) curve 2 has a broad flat minimum similar to that observed by Penning (1931) for neon contaminated with minute quantities ($\sim$0·002%) of argon, while the (V_s^-, pa) curve 1, on the other hand, is of the more usual Paschen form, similar to that obtained for a negative discharge in air and hydrogen. V_s^+ remained constant over a wide range of pa, and the (V_s^-, pa) and (V_s^+, pa) curves intersected at a pa of 1·13 mm Hg cm. After further outgassing for some hours V increased and curves 3 and 4 were obtained, and after prolonged heating *in vacuo* and severe ion bombardment the values given in curves 5 and 6 were obtained. V_s^- was now lower than V_s^+, but during the removal of occluded hydrogen both V_s^+ and V_s^- increased. From the (V_s, pa) data given in fig. 9 the coefficient ω/α as a function of X/p (fig. 10) was calculated using the values of α/p given by Townsend and MacCallum (1934). It is apparent that severe heating *in vacuo* and ion bombardment considerably decreased ω/α for the negative discharge, but produced only small changes in ω/α for the positive discharge.

Purification of the gas also had an important effect on the conditions under which the similarity theorem was obeyed. Although the heat treatment and ion bombardment increased both V_s^- and V_s^+ by a large factor (fig. 9), the theorem was obeyed only for a positive discharge (fig. 11). The difference between the values of aX_s for tubes B (black circles) and C (continuous curve) for a negative discharge was very large even after long and severe outgassing and gas purification treatment extending over some 60 hours. The outgassing was,

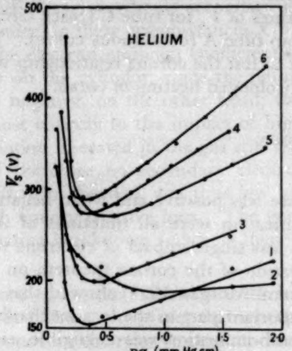


Fig. 9. Breakdown potentials of helium in various stages of purification for tube C. 1, 3 and 5 negative wire; 2, 4 and 6 positive wire.

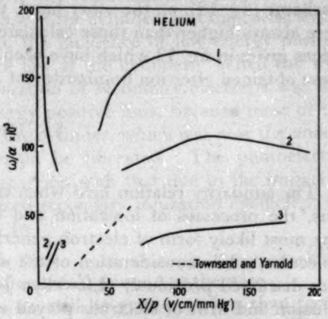


Fig. 10. $(\omega/\alpha, X/p)$ curves for nickel electrodes and helium in various stages of purification. 1, high concentration of atoms of impurity; 2, lower concentration of atoms of impurity; 3, helium in the final stages of purification.

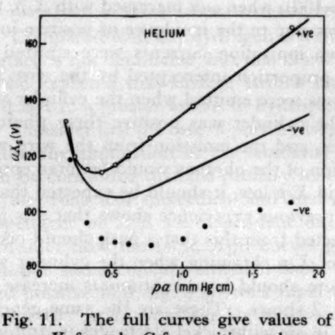


Fig. 11. The full curves give values of aX_s for tube C for positive and negative discharges. Points (white and black circles) give values of aX_s for positive and negative discharges in the geometrically similar tube B. The similarity theorem is obeyed for positive breakdown but not for negative.

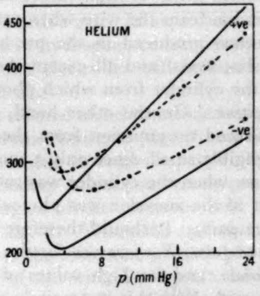


Fig. 12. The measured values (white and black circles) of V_s for tube C are compared with those (curves) calculated from the measured values for tube A using the corona relationship.

in fact, continued until sputtering of the electrodes eventually destroyed the insulation in the discharge tubes.

For a positive discharge (fig. 12) the measured values of V_s for tube C (white circles) were higher than those calculated from tube A (continuous curves) using the corona relationship (6) for values of p less than 9 mm Hg. For a negative discharge (fig. 12), on the other hand, the values of V_s for tube C (black circles) were always higher than those calculated from tube A (continuous curve). The values given in fig. 12 which have been used to test the corona relationship were those obtained after ion bombardment and prolonged heating *in vacuo*.

§4. DISCUSSION

(i) *Air*

The similarity relation held when the wire was positive and when negative: thus, the processes of ionization and de-ionization were all functions of X/p. The most likely form of electron generation was single impact of electrons with molecules, while consideration of the application of the corona theorem on the lines discussed previously (Llewellyn Jones and Morgan 1951) showed that the diffusion and drift of electrons played an important part in the loss mechanism.

For these relatively low values of pa photo-ionization was negligible, since the values of the photon absorption coefficient indicates that most of the photons would not be absorbed in the gas but would fall on the cylinder. It can be concluded (Llewellyn Jones 1939) from the form of the (ω/α, X/p) curves given in fig. 2 that at low values of X/p when ω/α decreased with increasing X/p the emission from the cylinder when cathode was mainly due to the incidence of photons, while, at high values of X/p (>100), when ω/α increased with X/p, the emission from the wire when cathode was due to the incidence of positive ions. Photons produced in the pre-breakdown ionization currents were emitted in all directions, and all except the small proportion intercepted by the wire fell on the cylinder from which photoelectrons were emitted when the cylinder was negative. On the other hand, when the cylinder was positive these photons produced no emission from the cylinder, and the emission from the wire was negligibly small, since only a small fraction of the photons would be intercepted. Thus, when the cylinder was cathode and X/p low, it should be expected that a part of the emission was photoelectric (previous experience shows that this is a large part). It should therefore be expected from this that a peak should occur in the (ω/α, X/p) curves for the values of X/p obtaining when the cylinder was cathode, and for high values of X/p there should be a continuous increase of ω/α with X/p; this is precisely what fig. 2 shows. These are the same general results as those observed for breakdown in air with nickel electrodes in uniform fields (Llewellyn Jones and Davies 1951 a); indeed, those previous values of ω/α practically bridge the gap between the two ranges of ω/α found in the present experiments in non-uniform fields. These results, therefore, are consistent with the view that the primary and secondary ionization processes leading to breakdown are the same in non-uniform fields as in uniform fields. The same secondary process of electron emission from the cathode due to incidence of photons and ions can and does occur, but the relative importance of the role they play in breakdown depends on the nature of the electrodes and the geometry of the system which determines the proportion of all the photons produced which fall on the cathode and the spatial distribution of electron and ion energies.

It follows from the above considerations that the generalized secondary coefficient ω/α for a positive wire may be approximately represented by the sum of two coefficients; thus from (4)

$$(\omega/\alpha)_+ = \gamma_l + (\delta/\alpha)_+, \qquad \ldots\ldots(7)$$

where $(\delta/\alpha)_+$ represents secondary emission due to incidence of photons on the cylinder (in these circumstances X/p was low at the cylinder) and the coefficient γ_l is the electron emission coefficient due to the incidence of low energy positive ions on the cylinder, since the fields were low at the cylinder. When the wire was negative, on the other hand, the liberation of secondary electrons was due almost entirely to the impact of high energy positive ions, because most of the radiation liberated in the gas still fell on the cylinder, which was now the anode, and therefore no secondary electrons could be liberated. The photoelectric yield from the wire was thus low in comparison with that due to the impact of high energy positive ions. The generalized secondary ionization coefficient for negative breakdown $(\omega/\alpha)_-$ is then given by

$$(\omega/\alpha)_- = \gamma_h, \qquad \ldots\ldots(8)$$

where γ_h was the value of γ at the high values of X/p found at the wire.

It follows, therefore, that whether $(\omega/\alpha)_-$ will be greater or less than $(\omega/\alpha)_+$ depends on the relative magnitudes of δ/α, γ_h and γ_l. Since ω/α is related to the breakdown potential V_s by expression (4), the relative magnitudes of the breakdown potentials for positive and negative breakdown depend on the relative magnitudes of δ/α, γ_h and γ_l, i.e. *on the nature of the cathode only*. The significance of this fact for corona breakdown has not been fully appreciated, as the discussion in § 1 shows.

(ii) *Hydrogen*

Compared with air, hydrogen is a simple non-active gas. In air a clean metal surface is not obtainable, but this is not the case in hydrogen; consequently it is to be expected that surface tarnish layers on the electrodes will influence V_s and ω/α to a greater extent in hydrogen than in air.

Consider first the case of pure hydrogen with cleaned nickel electrode surfaces. From the application of the similarity and corona theorems (figs. 6 and 7) it can be concluded that the primary ionization was due to single impact collision of electrons and molecules, and that the secondary ionization was a process dependent on X/p, i.e. dependent on the energies of the electrons and ions. Consider now the $(\omega/\alpha, X/p)$ curves given in fig. 8. When the cylinder was cathode and secondary emission could be due to the incidence of both photons and positive ions, fig. 8 shows a rapid increase of ω/α with X/p from a value of about 10^{-3} at 15 v/cm/mm Hg to a value of about 6×10^{-2} at 150 v/cm/mm Hg. When the wire was negative and photoelectron emission could not play any predominant role, ω/α steadily increased with X/p from very small values ($\sim10^{-3}$) at 150 v/cm/mm Hg to comparatively high values ($\sim6\times10^{-2}$) when X/p was $\sim$500 v/cm/mm Hg. This result can be interpreted as emission due to positive ions of hydrogen whose energy steadily increased with X/p at these high values of X/p.

It now remains to interpret the rapid increase of ω/α when X/p was less than 150 v/cm/mm Hg when the cylinder was the cathode. Clearly this could not be due to the incidence of positive ions, because when X/p was 150 v/cm/mm Hg and the wire negative, i.e. when photon emission was negligible, the value of ω/α due to positive ions alone was small ($\sim10^{-3}$). The rapid rise to such high values

($\sim 6 \times 10^{-2}$) when the cylinder was cathode and X/p was low must be interpreted as the influence of photons (Llewellyn Jones 1939); unfortunately calculation of ω/α for values of X/p greater than 150 v/cm/mm Hg with the cylinder as cathode was not possible as no reliable data of $\alpha/p = F(X/p)$ were available in order to solve eqn. (3). The dotted curve of fig. 8 gives the values of ω/α obtained with a cleaned nickel cathode in the experiments of Llewellyn Jones and Davies (1951 b) with uniform fields when about half the available photons could fall on the cathode. This dotted curve gives a rapid rise in ω/α from very small values when X/p was less than 150 v/cm/mm Hg to values of the same order but correspondingly smaller than those found when the cylinder was cathode (on which all the photons fell). While the agreement between the values of ω/α obtained in uniform fields with nickel and those obtained in the present work is not perfect, nevertheless, considering the fact that errors can be introduced in the solution of eqn. (3) using the published values of α, and also that different specimens of nickel were employed, the agreement can be regarded as very satisfactory. In comparing the full and dotted curves in fig. 8 it should be noted that in the case of non-uniform fields the curve at low values of X/p must be attributed almost entirely to the effect of photons, while the curve at the higher values of X/p (>150 v/cm/mm Hg) must be attributed to the effect of ions only (since it was electron emission from the wire on which only a few photons fell); for the case of uniform fields, on the other hand, the region between X/p of 150 v/cm/mm Hg (the photoelectric peak) and the low-pressure region when X/p is about 500, ω/α consisted of contributions from both the positive ion and the photon process (since about the same proportion ($\sim 50\%$) of the photons could always fall on the cathode).

The above conclusions are therefore consistent with the view that breakdown in pure hydrogen with clean surfaces in non-uniform fields involves the same ionization mechanisms as in uniform fields: the relative importance of the secondary process was, however, determined by the geometry of the systems.

Consider now the case when the electrodes were known to be covered with thin oxide layers. Previous work (Llewellyn Jones 1949) shows that the emission from a cathode due to incidence of positive ions is dependent on the nature and thickness of the surface film, because the film thickness determines the precise mechanism of electron extraction (Llewellyn Jones and de la Perrelle 1953, Llewellyn Jones and Morgan 1953) and so greatly influences the value of ω/α. The value of ω/α for a surface with a tarnish film may be greater or less than the value for a clean surface for a given value of X/p, depending on the thickness of the film. For thin tarnish layers ($\sim 10^{-7}$ cm obtained by a short exposure to air) the enhanced values of ω/α given by the top curve were obtained, while for comparatively thick films ($\sim 10^{-5}$ cm) the low values given by the bottom curve were obtained.

A property of corona breakdown with purified hydrogen, but with electrodes covered with oxide layers, was that at the lowest values of X/p (<35 v/cm/mm Hg) fig. 6 shows that the similarity relationship did not hold exactly for breakdown with a positive wire; for higher values of X/p, however, and also for all cases investigated when the wire was negative, the similarity relationship held within the experimental error. It is concluded therefore that the primary and secondary ionization processes were all functions of X/p, i.e. they depended on the electron and ion energies, except, however, for the case of the lowest value of

$X/p < 35$ v/cm/mm Hg for breakdown with a positive wire. The result indicates that, in this particular case, at least one process of generation or of loss was not a function of X/p. Traces of impurities in the hydrogen could not significantly affect the mechanism of generation, because breakdown in high-frequency fields when the generation was due to single impact collisions and the electrodes played no part (Llewellyn Jones and Williams 1953) showed that minute traces of impurity had no effect on the mechanism of generation or of loss because the similarity relationship was always obeyed. It is reasonable to assume, also, that processes of generation by electron impact and loss by diffusion and drift in the case of breakdown under steady fields were also processes dependent on X/p. It follows, therefore, that the important process which was not a function of X/p must have been a secondary ionization process which played an essential part in breakdown under uniform fields but played no part in breakdown under high-frequency fields. A probable process which can account for the phenomenon is that of enhanced emission from a cathode due to the high electric field set up by positive ions on thin surface insulating layers (Llewellyn Jones 1949, Llewellyn Jones and de la Perrelle 1953). The calculated values of ω/α are unfortunately too low to draw any reliable conclusions from them at values of X/p less than 35 v/cm/mm Hg. It is unlikely, however, that under these conditions the ion energies appreciably exceeded thermal values (Llewellyn Jones 1935, Tyndall 1938); also interaction with oxygen atoms on the oxide film cannot be ignored.

(iii) *Helium*

Since discharges in helium are extremely sensitive to the presence of small traces of impurities, the greatest care was taken in the present experiments to purify the gas and to clean and outgas the electrode surfaces. Nevertheless, it is not, of course, possible to say with certainty that no foreign molecules at all were present in the tubes. The results are therefore best discussed by considering the change in the characteristics of breakdown as the gas was progressively purified and the electrodes cleaned and degassed during the final stages of the purification when the proportion of impurities was minute. It should be emphasized that the electrodes were in a clean state after intensive ion bombardment and subsequent heating, but it should be remembered that to clean nickel completely is extremely difficult (Llewellyn Jones and Davies 1951 a, b). In the present case positive ion bombardment of the electrodes was continued until evaporation or sputtering of the metal eventually destroyed the insulation and rendered the tubes useless: the following results were obtained at various stages during this treatment.

At high values of X/p traces of impurities are not important (Kruithof and Penning 1937), and the values of α/p given for pure helium by Townsend and MacCallum (1934) are here used to calculate ω/α when the gas was in its final stages of purification: some extrapolation for values of X/p greater than 100 v/cm/mm Hg was necessary. Figure 10 gives the values of ω/α in terms of X/p thus calculated. The values of ω/α given by Townsend and Yarnold (1934) for helium are shown as a dotted curve in fig. 10, and these almost bridge the gap between the present values of ω/α for a positive and a negative wire. The $(\omega/\alpha, X/p)$ curves obtained in the present experiments have the same general features as those obtained for air and hydrogen in that, at low values of X/p when the cylinder was the cathode there appears to be a part of a photoelectric peak showing the preponderance of the photoelectric coefficient δ/α in the generalized coefficient ω/α. Further purification generally lowered the values of ω/α everywhere, but the sharp rise in ω/α with X/p at relatively low values of X/p when the wire was positive suggests that the values obtained when X/p was less than 20 v/cm/mm Hg were those corresponding to a photoelectric peak. On the other hand, when the wire was negative, and for high values of X/p greater than 50 v/cm/mm Hg, ω/α did not change in a marked manner with X/p; the general flatness of the curve suggests that the emission was almost independent of X/p.

An important feature of the curves is the considerable decrease in ω/α produced by the progressive removal of the final traces of impurities. For these measurements the electrode surfaces could be regarded as clean, especially the wire, which had suffered severe ion bombardment. When X/p was greater than 100 v/cm/mm Hg and the wire was the cathode, ω/α was almost entirely due to the positive ion γ effect. It is difficult, therefore, to account for the rapid fall in ω/α as the final traces of impurity were being removed in terms of the energies of helium positive ions. The fall in ω/α must consequently be interpreted as due to a change in the nature of the positive ions themselves as the impurity was progressively removed, since the surface conditions were constant.

Such a change can be produced by the action of metastable atoms which ionize atoms of the impurity in collisions of the second kind, so producing a considerable number of positive ions of the impurity molecules or atoms. Thus, when there were no impurities the positive ions were all those of helium, but when small traces of impurities were present the ions were mainly those of the impurity and their number proportional to the partial pressure of impurity. Three conclusions follow from this. Firstly, when the ions were those of helium there would be a comparatively small change in γ with increase of X/p, because exchange phenomena would prevent the helium ions acquiring high energies; Penning (1931) and Townsend and MacCallum (1934) have shown that γ changed only very slowly with X/p in pure monatomic gases. Secondly, with ions of impurities the converse was likely to hold. Owing to the large difference in the value of the ionization potential for the impurity and the ionization potential of the metastable level of helium, the probability of the occurrence of exchange phenomena was greatly reduced, with the result that the energy of the ions could more readily increase as X/p increased. Thirdly, if the ions were those of foreign atoms which had been produced by collisions of the second kind, their process of generation would not be in accordance with the similarity theorem; it should therefore follow that the similarity theorem would not hold when this type of process played a significant part in the breakdown process. These conclusions are precisely what the experimental results show. The similarity theorem did not hold for a negative wire, while, on the contrary, at low values of X/p and for the purest specimens of helium, the theorem held.

§5. CONCLUSIONS

The main conclusion which can be drawn from the results discussed in the present paper is that the electrical breakdown of gases at low pressures in non-uniform fields (a wire and a coaxial cylinder system) is consistent with the view that the breakdown mechanism is the result of primary and secondary processes of ionization. A number of secondary processes such as the emission due to incidence of positive ions or photons can and do occur, and the observed relative values of the breakdown potentials for positive and negative discharges are determined by the relative importance of the various possible secondary processes, which is itself determined by the polarity of the discharge. These cathode secondary processes can differ greatly in efficiency, so that the relative values of V_s for the positive and negative discharges therefore depend on the nature of the cathode surface. The mechanism of breakdown in non-uniform fields is thus of the same general nature as that for uniform fields for which it has been established that the relative importance of the secondary processes in any given gas is determined by the nature of the cathode surface. Further, a correlation has been established between the actual values of the secondary ionization coefficient for breakdown in non-uniform fields with those for uniform fields for the three types of gases, viz. air (a gas mixture), hydrogen (a pure diatomic gas) and helium (a monatomic gas).

ACKNOWLEDGMENTS

We wish to thank the Department of Scientific and Industrial Research for the award of a grant (1949–51) and the Council of the University of Wales for the award of a Fellowship to one of us (G. C. W.) to undertake this work. We also wish to thank Dr. G. D. Morgan and Mr. W. G. Townsend for helpful discussion.

REFERENCES

BOULIND, H. F., 1934, *Phil. Mag.*, **18,** 909.

BRUCE, J. H., 1930, *Phil. Mag.*, **10,** 476.

DUTTON, J., HAYDON, S. C., and LLEWELLYN JONES, F., 1953, *Proc. Roy. Soc.* A, in the press.

HALE, D. H., 1939, *Phys. Rev.*, **56,** 1199.

HASTED, J. B., 1951, *Proc. Roy. Soc.* A, **205,** 421.

HEALEY, R. H., and REED, J. W., 1941, *The Behavior of Slow Electrons in Gases* (Sydney : Amalgamated Wireless (Australasia)).

HUXLEY, L. G. H., 1928, *Phil. Mag.*, **5,** 721; 1930, *Ibid.*, **10,** 185.

HUXLEY, L. G. H., and BRUCE, J. H., 1937, *Phil. Mag.*, **23,** 1096.

KEENE, J. P., 1949, *Phil. Mag.*, **40,** 369.

KRUITHOF, A. A., and PENNING, F. M., 1937, *Physica*, **4,** 430.

LLEWELLYN JONES, F., 1935, *Proc. Phys. Soc.*, **47,** 74; 1939, *Phil. Mag.*, **28,** 192; 1949, *Proc. Phys. Soc.* B, **62,** 366.

LLEWELLYN JONES, F., and DAVIES, D. E., 1951 a, *Proc. Phys. Soc.* B, **64,** 397; 1951 b, *Ibid.*, **64,** 519.

LLEWELLYN JONES, F., and MORGAN, C. G., 1953, *Proc. Roy. Soc.* A, in the press.

LLEWELLYN JONES, F., and MORGAN, G. D., 1951, *Proc. Phys. Soc.* B, **64,** 560.

LLEWELLYN JONES, F., and DE LA PERRELLE, E. T., 1953, *Proc. Roy. Soc.* A, **216,** 267.

LLEWELLYN JONES, F., and WILLIAMS, G. C., 1953, *Proc. Phys. Soc.* B, **66,** 17.

LOUW, J. D., and NAUDE, S. M., 1949, *Phys. Rev.*, **76,** 571.

MILLER, C. G., and LOEB, L. B., 1951, *J. Appl. Phys.*, **22,** 494, 614, 740.

PENNING, F. M., 1931, *Phil. Mag.*, **11,** 961.

PIDD, R. W., and MADANSKY, L., 1949, *Phys. Rev.*, **75,** 1175.

SANDERS, F. H., 1933, *Phys. Rev.*, **44,** 1020.

TOWNSEND, J. S., 1915, *Electricity in Gases* (Oxford : Clarendon Press); 1947, *Electrons in Gases* (London : Hutchinson).

TOWNSEND, J. S., and MACCALLUM, S. P., 1934, *Phil. Mag.*, **17,** 678.

TOWNSEND, J. S., and YARNOLD, G. D., 1934, *Phil. Mag.*, **18,** 594.

TYNDALL, A. M., 1938, *The Mobility of Positive Ions in Gases* (Cambridge : University Press).

WIEDENBECK, M. L., and CRANE, H. R., 1949, *Phys. Rev.*, **75,** 1268.

The Viscosity of Gas Mixtures

By H. R. HEATH
Physics Department, Battersea Polytechnic

Communicated by T. L. Ibbs; MS. received 10th February 1953

Abstract. The viscosity of binary mixtures over the complete range of concentration has been measured in the case of He–A, He–N_2, He–CO_2 and H_2–A, H_2–N_2, H_2–CO_2. All the measurements were made at 18°c and 70 cm of mercury pressure by a modified form of Rankine viscometer. The results are in very good agreement with the values calculated from Chapman's theory. Values for the first approximation to the diffusion coefficient $(D_{12})_1$ are obtained from the results, and these are in approximate agreement with measured values where the latter are available.

---

§ 1. INTRODUCTION

CHAPMAN (1915, 1916a, b) and Enskog (1917) have independently made theoretical investigations of the viscosity of mixtures of gases. Formulae have been obtained for the viscosity of a binary mixture, the molecules of which during interactions obey a law of force of the type $f(r) = Kr^{-\nu}$, where r is the distance apart and ν the repulsive force index. The theory is given in a compact form by Chapman and Cowling (1939).

In order to examine the theoretical predictions experiments have been made to measure the viscosity over the whole range of concentration for mixtures of helium–argon, helium–nitrogen and helium–carbon dioxide, i.e. a monatomic gas mixed with a monatomic gas, a diatomic gas and a polyatomic gas. Measurements have likewise been made on mixtures of hydrogen–argon, hydrogen–nitrogen and hydrogen–carbon dioxide, i.e. a diatomic gas mixed with a monatomic gas, a diatomic gas and a polyatomic gas. The temperature and pressure of the mixtures were maintained constant at 18°c and 70 cm of mercury respectively throughout the series of experiments.

§ 2. EXPERIMENTAL

In order to obtain results for a large number of gas mixtures the Rankine (1910) viscometer was used in the modified form shown in fig. 1. It was made in hard glass and attached to a gas burette and pump via a conical ground joint J. With vacuum grease on the joint the viscometer could be rotated about J as axis and the viscosity of the gas measured by observing the time of fall of the mercury pellet between two fixed marks engraved on the tube B. Once the joint at J had been made there was no need to remove the viscometer in the whole series of measurements. The capillary tube A varied in diameter between 0·24 mm at one end and 0·30 mm at the other, but as the apparatus was standardized by the use of a gas of known viscosity this variation does not affect the results. The correction for 'slip' will be negligible for all the gases used. The tube B had a uniform diameter of 2·40 mm. The overall length of the tubes was 50 cm and the width 12 cm.

In order to evacuate the viscometer it was first turned at an angle of 45° to the vertical so that the mercury could run down and form a globule in the corner of the rectangle. The taps were opened and the pressure reduced to about 10^{-3} cm of mercury by means of a Töpler pump. The pump was now isolated and the gas admitted to the viscometer from a gas burette, and its pressure adjusted to 70 cm of mercury.

The experiments were made in a small room and the temperature of the whole room was maintained at 18°c by the use of electric heaters. A thermometer reading to 0·1 centigrade degree was attached to tube A (fig. 1) and measurements were only made when it was quite steady and reading $18{\cdot}0 \mp 0{\cdot}2$°c.

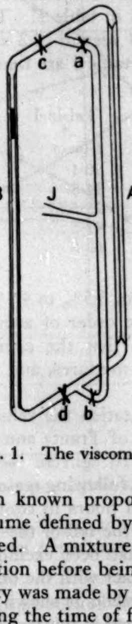


Fig. 1. The viscometer.

Gas mixtures were made in known proportions by volume by measuring partial pressures in a fixed volume defined by an index to which the mercury level in the burette could be raised. A mixture was left for a period of two hours to attain uniformity of composition before being admitted to the viscometer.

The measurement of viscosity was made by closing taps a and b, setting the viscometer vertical and measuring the time of fall of the mercury pellet between the two fixed marks on B (fig. 1), when taps c and d were opened. The distance between the marks was 35 cm.

Purity of the Gases

With the exception of helium, the gases used were obtained in cylinders, and the manufacturers estimate the impurity generally as not exceeding 0·2%. The helium was obtained in a sealed glass flask and was spectroscopically pure.

The Capillarity Effect

The driving pressure of the mercury pellet is reduced by the effect of surface tension. The fractional value of this reduction can be obtained by the method used by Rankine and Smith (1921). The mercury pellet is first timed in the usual way and the measurement is then repeated with the pellet broken into two approximately equal segments a small distance apart. Let the fractional reduction for the single pellet be x and the time of fall t_1. For the divided pellet the fractional reduction will be $2x$, and let the time of fall be t_2. Then $(1-x)t_1=(1-2x)t_2$, giving $x=(t_2-t_1)/(2t_2-t_1)$. Rankine and Smith report a variation in the value of x both with the temperature and the nature of the gas. The present experiments were all made at constant temperature, and using gases much less likely to produce any chemical effect on the surface of the mercury pellet than those used by Rankine and Smith. Measurements of x were made in the manner described for all the pure gases to be used. Four observations of each time of fall were made and the mean is recorded in table 1. The maximum variation between extremes in these times was 0·4 second (0·2 second in the case of hydrogen). The values of t_1 and t_2 given in table 1 are therefore probably correct to within 0·5%. The values of x vary from 3·5% to 4·3%, i.e. they agree with the mean value of 3·9% to about the same order of accuracy as the measurements of t. It has therefore been assumed that the capillarity correction was constant throughout a particular series of mixtures, and the viscosity proportional to the time of fall of the single pellet.

Table 1

Gas	t_1 (sec)	t_2 (sec)	x
Nitrogen	50·1	52·4	0·042
Helium	54·8	57·1	0·039
Hydrogen	25·5(5)	26·5	0·035
Carbon dioxide	41·5	43·4(5)	0·043
Argon	62·8	65·2	0·036

For the purpose of standardization the value of the viscosity of nitrogen at 18°c was obtained from the work of Trautz and Baumann (1929) and of Rigden (1938) and found to be $174{\cdot}0\times10^{-6}$ gm cm^{-1} sec^{-1}. Air was not used for the purpose of standardization for the following reasons. It was found that when air, which had been dried for several hours in contact with phosphorus pentoxide, was admitted to the viscometer, the fall of the pellet was not perfectly regular. It was also found that when air had been excluded from the apparatus for some time the fall of the pellet in contact with the other gases was quite smooth and regular. The validity of the method is shown by the agreement between the relative values of the results obtained in this way for the pure gases and the results for the same gases of other observers using absolute methods.

§ 3. RESULTS

The results obtained are shown as the experimental points in fig. 2. The curves drawn in fig. 2 are values of $[\mu]_1$, the first approximation to μ for mixtures calculated from the results of Chapman's theory.

Experimental values of μ for the pure gases are in close agreement with the results of Trautz and his collaborators (1929, 1930). The agreement between the theoretical and experimental curves is very close, although the shape is very different for different gas pairs.

Graham (1846) first noticed the curious fact that the addition of a limited amount of light and less viscous gas (hydrogen in the case mentioned) to a heavy and more viscous gas (carbon dioxide) may increase the viscosity. This is even more noticeable in the case of helium and argon, where the values of the viscosity of the pure constituents are closer together. The viscosity of a gas, unlike its

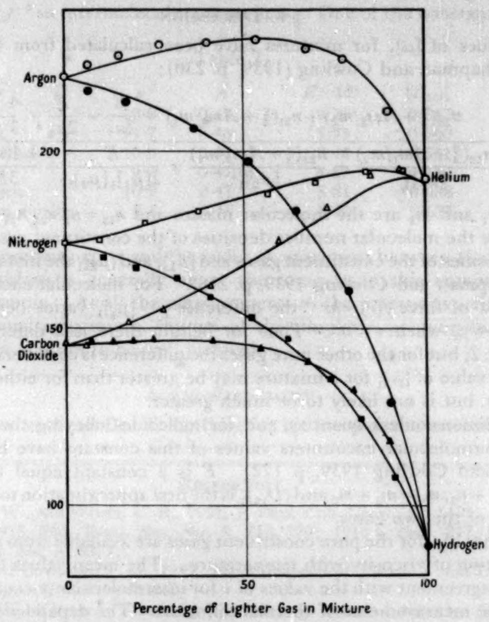


Fig. 2. The viscosity of some gas mixtures at 18°C. Experimental results shown by points; the curves drawn show calculated values.

thermal conductivity, is not closely related to its molecular weight. The light gas helium has a viscosity greater than that of carbon dioxide or nitrogen. From fig. 2 it appears that the addition of a small amount of heavy but less viscous gas (carbon dioxide or nitrogen) to a light but viscous gas (helium) may also increase the viscosity.

Mixtures of helium and argon have also been examined by Trautz and his collaborators (1929, 1930), and Chapman and Cowling (1939, p. 233) have shown that their results are in close agreement with theory. The range of concentration in their experiments was, however, rather limited. In particular there were no mixtures used containing less than 35% of helium and, as will be seen, it is in this range that a small difference between the theoretical and experimental curves may exist.

A number of binary gas mixtures have also been examined more recently by Buddenberg and Wilke (1951), including hydrogen–carbon dioxide mixtures at a temperature of 25°C. Allowing for the temperature difference, the results for the three concentrations they used are in close agreement with those shown in fig. 2 for mixtures of this pair of gases.

§ 4. DISCUSSION

The values of $[\mu]_1$ for mixtures have been calculated from the expression given by Chapman and Cowling (1939, p. 230):

$$[\mu]_1 = \frac{n_{12}(\frac{2}{3} + Am_1/m_2) + n_{21}(\frac{2}{3} + Am_2/m_1) + \dfrac{E}{2\mu_1} + \dfrac{E}{2\mu_2} + \dfrac{4}{3} - 2A}{\dfrac{n_{12}(\frac{2}{3} + Am_1/m_2)}{[\mu_1]_1} + \dfrac{n_{21}(\frac{2}{3} + Am_2/m_1)}{[\mu_2]_1} + \dfrac{E}{2[\mu_1]_1[\mu_2]_1} + \dfrac{4A(m_1 + m_2)^2}{3E\,m_1 m_2}},$$

in which m_1 and m_2 are the molecular masses and $n_{12} = n_1/n_2$, $n_{21} = n_2/n_1$, where n_1 and n_2 are the molecular number densities of the constituent gases. μ_1 and μ_2 are the viscosities of the constituent gases and $[\mu_1]_1$ and $[\mu_2]_1$ the first approximation values (Chapman and Cowling 1939, p. 218). For molecular encounters which obey the law of force $f(r) = kr^{-\nu}$ the difference $\mu - [\mu_1]_1$ varies between 0 when $\nu = 5$ and 1·6% when $\nu = \infty$. Thus for helium there is a difference of 0·8% shown in fig. 2, but for the other pure gases the difference is considerably less. The error in the value of $[\mu]_1$ for a mixture may be greater than for either of the pure constituents, but is not likely to be much greater.

A is a dimensionless quantity, and for molecules obeying the above law of force in intermolecular encounters values of this constant have been tabulated (Chapman and Cowling 1939, p. 172). E is a constant equal to $\frac{2}{3}nm_0(D_{12})_1$, where $n = n_1 + n_2$, $m_0 = m_1 + m_2$ and $(D_{12})_1$ is the first approximation to the coefficient of diffusion of the two gases.

The values of ν for the pure constituent gases are available from measurements of the variation of viscosity with temperature. The mean values for various gas pairs are in agreement with the values of ν for intermolecular encounters obtained directly from measurements of thermal diffusion. The dependence of A on the value of ν is not very marked, so that an error in the value of ν will not seriously affect the value for A. The quantity E has been chosen to fit the experimental observations for a fifty per cent mixture (fig. 2). The theoretical curves may thus be expected to fit closely at the extremes and exactly in the middle.

It will be seen that there is very close agreement throughout between the theoretical and experimental curves over the whole range of mixtures examined. There are small departures from an exact fit for mixtures of helium and argon containing less than 30% of helium, and for mixtures of helium and carbon dioxide containing more than 60% of helium. In these two cases there is however close agreement between the general shapes of the curves. The results of these experiments thus give further support to the correctness of the theoretical conclusions.

Table 2 gives the values of ν, A and E which have been used in the calculation of $[\mu]_1$, together with values of $(D_{12})_1$ at a temperature of 18°C and a pressure 76 cm of mercury which have been obtained from the values of E. The last column has been included to show values of D_{12} (experimental) where these are available. The last two columns are in decidedly good agreement with the exception of the values for hydrogen–nitrogen. The difference in this case may be partly error in the approximation and partly experimental error.

Buddenberg and Wilke (1951) have preferred to compare their results with the theoretical relations of Hirschfelder, Bird and Spotz (1948, 1949) which have been obtained from an exact solution for molecules obeying the law $f(r)=k_1r^{-13}-k_2r^{-7}$ in intermolecular encounters. One of the constants is chosen to fit as in the present paper, and the order of agreement between theoretical and experimental values is generally about the same as in this comparison with Chapman's formula. As all the measurements in the present experiments were made at the same temperature, the use of the simpler expression for the intermolecular force appears to be well justified.

Table 2

Gas pair	ν	A	$E\times10^4$	$(D_{12})_1$	$(D_{12})_{\mathrm{exp}}$
Helium–argon	10·0	0·47	9·20	0·742	0·71(6)
Helium–nitrogen	11·7	0·46	5·95	0·660	—
Helium–carbon dioxide	10·0	0·47	7·48	0·551	—
Hydrogen–argon	9·3	0·47(6)	8·87	0·750	0·77
Hydrogen–nitrogen	10·0	0·47	5·81	0·688	0·75
Hydrogen–carbon dioxide	8·5	0·48	8·20	0·633	0·61(6)

ACKNOWLEDGMENT

I would like to thank Dr. T. L. Ibbs for his interest and encouragement in this work.

REFERENCES

BUDDENBERG, J. W., and WILKE, C. R., 1951, *J. Phys. Coll. Chem.*, **55,** 1491.

CHAPMAN, S., 1915, *Phil. Trans. Roy. Soc.* A, **216,** 279; 1916 a, *Ibid.*, **217,** 115; 1916 b, *Proc. Roy. Soc.* A, **93,** 1.

CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-Uniform Gases* (Cambridge: University Press).

ENSKOG, D., 1917, *Kinetische Theorie der Vorgänge in massig verdünnten Gasen*, Inaug. Dissertation, Upsala.

GRAHAM, T., 1846, *Phil. Trans. Roy. Soc.*, **136,** 573.

HIRSCHFELDER, J. O., BIRD, R. B., and SPOTZ, E. C., 1948, *J. Chem. Phys.*, **16,** 968; 1949, *Chem. Rev.*, **44,** 205.

RANKINE, A. O., 1910, *Proc. Roy. Soc.* A, **83,** 265.

RANKINE, A. O., and SMITH, C. J., 1921, *Phil. Mag.*, **42,** 603.

RIGDEN, P. J., 1938, *Phil. Mag.*, **25,** 961.

TRAUTZ, M., and BAUMANN, P. B., 1929, *Ann. Phys. Lpz.*, **2,** 733.

TRAUTZ, M., and BINKELE, H. E., 1930, *Ann. Phys. Lpz.*, **5,** 561.

TRAUTZ, M., and KIPPHAN, K. F., 1929, *Ann. Phys. Lpz.*, **2,** 743.

Relation between Velocity of Sound and Viscosity in Liquids

By S. PARTHASARATHY and N. N. BAKHSHI
National Physical Laboratory of India, New Delhi

Communicated by E. N. da C. Andrade; MS. received 29th September 1952

Abstract. A new relationship, viz. $v^{1/3}/\rho = A + B/\eta^{1/2}$, between sound velocity v, viscosity η and density ρ of a liquid, has been obtained. It is obskrved that A is a constant ($=13{\cdot}56$) for all the homologous series considered, whereas B is different for different series. A plot of $v^{1/3}/\rho$ against $1/\eta^{1/2}$ gives a set of straight lines diverging from the same point on the y-axis (0, 13·56).

---

§ 1. INTRODUCTION

No extensive data existed on sound velocities in organic liquids until about 1936–37, and therefore no correlation between them and the structure of the compounds was possible. At about this time one of us (Parthasarathy 1938) derived several empirical relationships regarding sound velocity and chemical constitution from the measurements of sound velocity in organic liquids. Subsequently this evoked interest in the subject, and as examples of theoretical study the works of Rao (1940), Lagemann and Dunbar (1945), Schaaffs (see Bergmann 1949), Parshad (1945) and others can be mentioned.

In this paper it is proposed to examine whether any relationship exists between sound velocity and viscosity.

§ 2. RESULTS AND DISCUSSIONS

Parthasarathy (1935) has already shown that, in general, apart from a few exceptions in which heavy atoms are substituted, lengthening of the molecular chain enhances the sound velocity. A further development of this idea has been expressed by Rao (1940) in a formula $v^{1/3} M/\rho = R$ where v is the velocity of sound, M the molecular weight, ρ the density and R a constant which is independent of temperature for a given series. In a paper to be published the authors have shown that the above formula could be expressed with advantage as $v^{1/3}/\rho = A + B/M$ where A and B are constants.

It is well known from the work of Gartenmeister (1890) that η/M^2 is a physico-chemical constant, η being the viscosity, at all temperatures, for all series in which successive members differ by CH_2. There are, therefore, two relations, one between the sound velocity and the molecular weight and the other between the molecular weight and viscosity. By eliminating M between the two a relationship between sound velocity and viscosity can be derived.

The new expression obtained in this manner is of the form $v^{1/3}/\rho = A + B/\eta^{1/2}$. This is analogous to that obtained in the case of molecular weight. Therefore it can be predicted that if $1/\eta^{1/2}$ is plotted against $v^{1/3}/\rho$ a set of straight lines should be obtained all diverging from the same point. This is amply confirmed as shown by fig. 2. The different series investigated in the present work are tabulated (see table). The values for the viscosity of the organic liquids

Liquid	$v^{1/3}/\rho$	η_{20}	$\eta^{1/2}\times 10^2$	$1/\eta^{1/2}$
Aliphatic Hydrocarbons.				
Pentane	16·02	0·00227	4·76	21
Hexane	15·57	0·00318	5·64	17·73
Heptane	15·38	0·00411	6·41	15·6
Octane	15·11	0·00542	7·36	13·6
Nonane	15·02	0·00711	8·43	11·86
Fatty Acids.				
Propionic acid	10·64	0·01099	10·48	9·52
Butyric acid	11·09	0·01538	12·4	8·06
Valeric acid	11·41	0·02300	15·17	6·59
Caproic acid	11·69	0·03198	17·88	5·59
Aliphatic Alcohols.				
Methyl alcohol	13·14	0·00593	7·70	12·99
Ethyl alcohol	13·39	0·01192	10·91	9·17
Propyl alcohol	13·30	0·02255	15·02	6·66
Butyl alcohol	13·37	0·02947	17·16	5·83
Esters.				
Ethyl formate	10·86	0·00398	6·31	15·84
Ethyl acetate	11·73	0·00450	6·71	14·91
Ethyl propionate	11·97	0·00531	7·29	13·73
Ethyl butyrate	12·01	0·00668	8·17	12·24
Alkyl Iodides.				
Methyl iodide	4·13	0·00487	6·98	14·33
Ethyl iodide	4·92	0·00583	7·63	13·10
Propyl iodide	5·58	0·00737	8·58	11·65
Ethers.				
Diethyl ether	14·04	0·002448	4·95	20·21
Dipropyl ether	13·87	0·00425	6·52	15·34

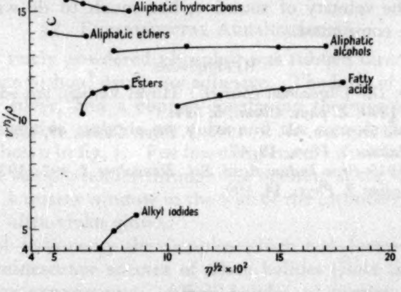


FIG. 1.

have been taken from the *International Critical Tables* and the *Landolt–Börnstein physikalisch chemische Tabellen*. The sound velocity values have been reduced to 20°C, as in the case of η.

It is observed from fig. 1 that the curves of $v^{1/3}/\rho$ plotted against $\eta^{1/2} \times 10^2$ show a marked variation in slope. Starting from aliphatic hydrocarbons it is found that the slope decreases in passing from aliphatic hydrocarbons to ethers, for alcohols it is almost a straight line parallel to the x-axis; then the slope changes its sign and increases negatively in the order esters, aliphatic acids and alkyl iodides. On the other hand, if graphs of $v^{1/3}/\rho$ against $1/\eta^{1/2}$ are examined they are all found to be straight lines and, remarkably enough, diverging from the same point, viz. (0, 13·56). This means, as the factor $1/\eta^{1/2}$ decreases or $\eta^{1/2}$ increases, the factor $v^{1/3}/\rho$ tends to 13·56 in all cases.

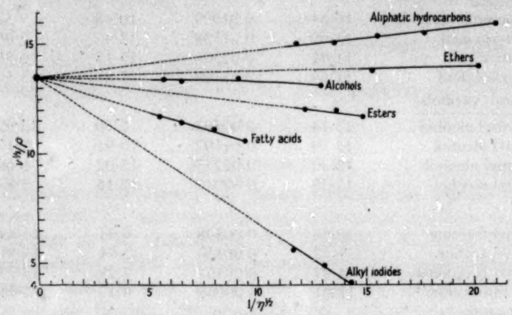


FIG. 2.

This in a way confirms Gartenmeister's finding (1890) that η/M^2 is a constant for all series in which successive members differ by CH_2. The point (0, 13·56) on the y-axis, from which all the straight lines of the graph ($v^{1/3}/\rho$, $1/\eta^{1/2}$) diverge, seems to be of fundamental importance, though little is known about it at present.

The marked variation of the slope, as stated above, suggests that the factor $v^{1/3}/\rho$, or rather the velocity of sound v, has much to do with the molecular constitution of the compound.

REFERENCES

BERGMANN, L., 1949, *Der Ultraschall* (Zurich : Hirzel Verlag), 5th edn.
GARTENMEISTER, R., 1890, *Z. phys. Chem.*, **6,** 524.
LAGEMANN, R. T., and DUNBAR, W. S., 1945, *J. Phys. Chem.*, **49,** 428.
PARSHAD, R., 1945, *Indian J. Phys.*, **19,** 47.
PARTHASARATHY, S., 1935, *Proc. Indian Acad. Sci., Bangalore*, **2,** 497; 1938, *Curr. Sci.*, **6,** 322.
RAO, M. R., 1940, *Indian J. Phys.*, **14,** 109.

Decay of Long Period Afterglow of Alkali Halides under Cathode Ray Excitation

By H. N. BOSE and J. SHARMA
Khaira Laboratory of Physics, University College of Science, Calcutta

Communicated by H. W. B. Skinner; MS. received 5th November 1951, *and in amended form 27th October* 1952

Abstract. The long period afterglow of pure and activated alkali halides excited with cathode rays has been studied at room and liquid oxygen temperatures. It has been found that the decay follows a power law $I \propto t^{-n}$ between 10 and 100 seconds. The value of the power constant n lies in general between 0·7 and 1·1 and is dependent on the exciting intensity. The decay rates are different for different parts of the spectrum. An attempt has been made to explain the results by an extension of the existing theory developed for photoluminescence.

§1. INTRODUCTION

In recent years there has been a great deal of experimental and theoretical work on the luminescence of pure and activated crystals under ultra-violet excitation; the temperature dependence of long period decay curves has been satisfactorily explained in many cases by assuming the presence of trapping centres at various depths below the vacant conduction band in the crystalline material. The present work reports on decay of afterglow of pure and activated alkali halides bombarded by cathode rays (at different temperatures) and attempts to explain the results obtained by an extension of the theory developed for photoluminescence (Randall and Wilkins 1945, Garlick and Gibson 1947, etc.).

§2. EXPERIMENTAL ARRANGEMENTS

A thin layer of finely powdered phosphor was rubbed directly on the surface of the sample holder without using any adhesive. The bulb of the sample holder was made of thin silver, and a copper–constantan thermocouple was soldered to it. The demountable cathode-ray tube and the sample holder used in the present work are shown in fig. 1. For low-temperature work liquid oxygen could be poured into the sample holder through the opening at the top which projects outside the tube. A quartz window in the wall of the cathode-ray tube permitted observation of the ultra-violet glow.

For the sake of uniformity the samples which had formerly been used for photographing luminescence spectra of alkali halides (Bose and Sharma 1950) were used in these experiments. Alkali halides of reagent quality procured from E. Merck were used as pure samples; activated mixtures were obtained by melting a measured quantity of impurity together with the pure reagent in an electric furnace where the molten state was maintained for a fairly long time to ensure homogeneity of the sample. The intensity measurements have been carried out in a darkened chamber with R.C.A. photomultiplier tubes 1P22, 931A, 1P28. The tubes were operated from a valve-regulated power-unit of conventional type; accelerating voltages per dynode stage were maintained between 70 and 80 v and the output current was always kept below 100 μA. Adjustment was checked for linear response of the photocell used and dark current was kept less than 0·5 μA.

The decay of afterglow intensity was followed by visual observations of the readings of a standard microammeter and by photographing on a rotating drum the deflections from a critically damped galvanometer. The effect of damping on the time-lag of galvanometer deflections was investigated and it was found that 5 seconds after start the correction does not appreciably change the form of the curve. The temperature was recorded on the same drum with the help of another Moll galvanometer connected to the thermocouple.

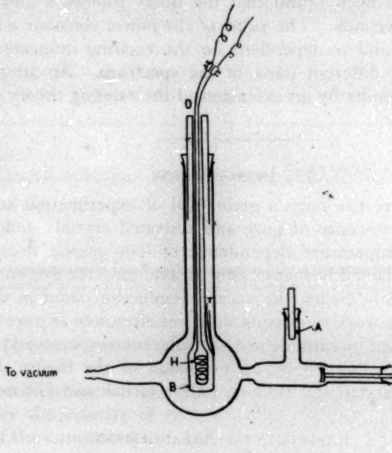


Fig. 1. Diagrammatic representation of the demountable cathode-ray tube used in the present work. A, anode, earth-connected; B, silver bulb of the sample holder; C, cathode; H, heater; O, opening for pouring liquid oxygen; Q, quartz window; T, thermocouple.

To obtain the decay rates in different spectral regions during the same decay experiment two suitable photomultipliers with necessary filters were placed side by side in front of the sample. The output of the two photocells was fed to the two galvanometers and thereby a simultaneous record was secured.

§3. Results and Discussions

The decay of afterglow has been investigated for pure and activated NaCl, NaBr, KCl, KBr and KI. It will be seen from records that during the initial and final stages the decay curves are different from those obtained during the intermediate stage between 10–100 seconds. The present work centres about this intermediate decay region, where, in almost all cases, the curve (log I, log t) is a straight line. The slopes of the curves have been determined for the above-mentioned samples and in different spectral regions of the afterglow. Some of the results are given in the table. During the initial stage of the decay the nature of the curve obtained is entirely different and the decay rate is extremely high; no attempt has been made to follow this part in any detail. After about two minutes the decay rate slows down enormously and the record curve now runs almost parallel to the axis, so that measurements become uncertain. Our observations have been mainly concentrated on the afterglow decay between 10 and 100 seconds. The decay constant is greatly dependent on the intensity of excitation (the value of n increasing with an increase in excitation intensity) and the decay curves are greatly modified by changes in excitation intensity. As comparison between two curves taken separately is thus not justified, all reported comparisons of decay curves in different spectral regions of the same sample were made from simultaneous records.

Sample	Temp. (°K)	Photocell	Spectral region (Å)	Slope	Remarks
NaCl	90	1P28	4000–3000	0·76	Recorded simultaneously
	90	931A	Visible: red to 4200	0·86	
	300	931A	Visible	0·72	
	465	931A	Visible	0·95	
NaCl : Tl 2·5%	90	1P28	Visible and u. v.	0·91	Recorded simultaneously
	90	931A	Visible	0·85	
KCl	90	931A	Visible	0·64	Depending on excitation
				0·71	
KCl : Tl 2·5%	90	931A	Visible	0·6	Excitation intensity kept low and constant
	100	931A	Visible	0·7	
	194	931A	Visible	1·0	
	307	931A	Visible	1·0–1·1	
	340	931A	Visible	1·1–1·6	
	90	931A	Visible	0·86	Recorded simultaneously
	90	1P28	4000–3000	0·96	
KBr : Tl	90	931A	Visible	0·74	
	308	931A	Visible	1·07	

For pure alkali halides the emission can be broadly grouped into four regions with respect to their decay rates: (*a*) red region of wavelengths longer than 6040 Å, (*b*) yellow-green region between 6000 and 5100 Å, (*c*) blue between 5100 and 4000 Å, and (*d*) emissions shorter than 4000 Å. The red emission has not been observed in the afterglow as the decay in this region is very rapid. The decay rate diminishes from region (*b*) to region (*c*), while in the ultra-violet region the rate is still smaller. According to modern conceptions the decay is due to the time spent by the electron in the trapping centre, the recombination of the released electron with the emitting centre requiring time of smaller order. Therefore the different decay constants for different spectral regions clearly show that the emissions in the different spectral regions are not due to electrons released from the trap of same depth. The electrons from a particular type of trap have a preference for centres which are responsible for emissions in particular regions. Thus at least some of the emitting centres are closely associated with the trapping centres.

In thallium activated samples the decay of red emission is, as in pure samples, too quick for our recording system. The ultra-violet part of the afterglow emission however decays now at a more rapid rate than does the visible part. Thus the thallium band which occurs near the ultra-violet region has a more rapid rate of decay than the visible emission bands, which are assigned to luminescence of matrix. Separate experiments on thermoluminescence of these specimens have shown that the traps associated with the thallium centres are deeper than those associated with the rest of the emission. As such, the decay of afterglow emission was expected to be slower for the thallium bands. The observed quicker decay indicates that stored energy in trapped Tl centres is being transferred to other centres by some non-radiating process. This may excite levels operative during luminescence of pure samples, so that there is increased emission of the bands of the matrix. This conclusion is also supported by the spectroscopic study of cathode luminescence of these samples (Bose and Sharma 1950). It was found that the presence of an activating impurity not only creates new emitting centres but also enhances the host crystal emissions. The photo excitation of the phosphor in the absorption band of the impurity produces the characteristic impurity as well as the host crystal emissions.

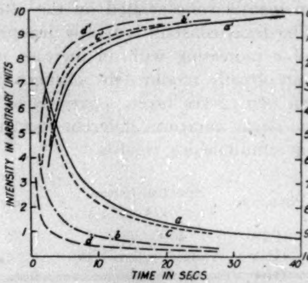


Fig. 2. Simultaneous records of the afterglow of KCl : Tl 2·5% at 90°K, with two galvanometers connected to different photomultipliers. The curves marked with the same letter give decay in different parts of the spectrum during the same experiment. The two galvanometers have their movements in opposite directions.

a. 1P28 for total emissions across quartz.
b. 1P28 with Wood's filter.
c. 1P28 with blue filter.
d. 1P28 with yellow filter.
a′, *b′*, *c′*, *d′* are corresponding curves with 931A.

The mechanical mixture of NaCl + TlCl behaves in a manner similar to pure NaCl; the mechanical mixture of KCl + TlCl is known to yield a luminescence spectrum almost identical with that of KCl : Tl (prepared from a melt) (Bose and Sharma 1950). In KCl + TlCl the afterglow intensity for the characteristic thallium band is very poor. The difference in decay rates for different spectral regions is similar to that in the case of KCl : Tl.

Garlick (1949) has shown that, in the case of retrapping, a power law $I \propto t^{-2}$ can be expected at long decay times for the photoluminescence of phosphors having a single trapping mechanism of sharp depth. The non-saturation of the traps due to low excitation intensity should make the afterglow intensity decay rather slowly at the beginning; the log–log plot of the decay curves, obtained for different excitation intensities, should approach a common asymptote with a slope of -2 when decay is followed for longer periods.

The behaviour of the decay curves of this report is different from the above predictions of the theory probably for two reasons: (i) these phosphors have numerous traps of different depths, (ii) excitation by low-energy cathode rays cannot bring into operation traps throughout the entire volume of the specimen. The release of electrons from the traps whose glow temperatures are below the temperature of excitation takes place very quickly, so that the contribution of such trapping mechanisms to the long period afterglow is negligible. The energy stored up in the deeper traps is also very small when the phosphor is excited at a temperature much below the corresponding glow temperatures. Thus the long period decay curves which have been studied in the present investigations are mainly determined by the electron traps whose glow temperatures lie near the temperature of excitation. The deeper traps are responsible for the tail end of the decay curves which is observed to have a very small decay constant. Therefore the decay constant cannot be expected to vary directly with temperature: rather the pattern of glow peaks should be reflected in the variation. This is borne out by the behaviour of NaCl. The decay constant of NaCl has a high value at liquid oxygen temperature, its value becomes small at room temperature and again becomes great at higher temperatures. This is supported by the fact that sodium chloride gives an isolated glow peak at low temperature and the next peak occurs at 575°K. At the intermediate temperatures, far from the peaks, the decay constant becomes small. This

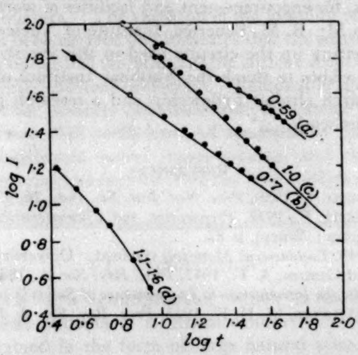


Fig. 3. Log I plotted against log t for afterglow decay of KCl : Tl 2·5% at different temperatures. (*a*) 90°K, (*b*) 100°K, (*c*) 307°K, (*d*) 340°K.

behaviour is different in KCl, for which only one strong glow peak has been located at low temperature. In KCl : Tl also this behaviour is not clear because the sample shows a number of glow peaks occurring close together at low temperatures (to be reported elsewhere).

The low value of power law constant n can be ascribed to a number of factors: firstly some of the traps may be remaining unsaturated ; secondly there exist more than one trap of different depths and these contribute simultaneously to the afterglow; lastly the electrons from the deeper traps have a high probability of being retrapped in shallow traps, which have glow temperature near the temperature of excitation. Naturally these effects will be more prominent at higher temperatures where the (log I, log t) curve does not remain a straight line. This is seen in the decay curve of KCl : Tl at 340°K, shown in fig. 3.

ACKNOWLEDGMENTS

Our thanks are due to Professor S. N. Bose, Khaira Professor of Physics, Calcutta University, for encouragement and facilities of work. We acknowledge our indebtedness to Mr. B. M. Banerjee, Institute of Nuclear Physics, Calcutta, for helping us in setting up the circuits used in this investigation. One of the authors (H. N. B.) wishes to thank the National Institute of Sciences of India, for the I.C.I. Research (India) Fellowship and a research grant, which helped him to carry out this work.

REFERENCES

BOSE, H. N., and SHARMA, J., 1950, *Proc. Nat. Inst. Sci. Ind.*, **26,** 47.

FONDA, G. R., and SEITZ, F., 1948, *Preparation and Characteristics of Solid Luminescent Materials* (London : Wiley), p. 87.

GARLICK, G. F. J., 1949, *Luminescent Materials* (Oxford : University Press), p. 37.

GARLICK, G. F. J., and GIBSON, A. F., 1947, *Proc. Roy. Soc.* A, **188,** 485.

LEVERENZ, H. W., 1950, *An Introduction to Luminescence of Solids* (London : Wiley), p. 292.

RANDALL, J. T., and WILKINS, M. H. F., 1945, *Proc. Roy. Soc.* A, **184,** 365.

Electrical Properties of Molybdenite

By R. MANSFIELD and S. A. SALAM
Bedford College, University of London

Communicated by H. T. Flint; MS. received 19*th September* 1952, *and in final form* 22*nd December* 1952

Abstract. The following electrical properties of natural crystals of MoS_2 have been measured: conductivity, Hall coefficient and thermoelectric power over the temperature range -183°c to 500°c, and the room temperature change of conductivity in a magnetic field. It was found that the majority of the specimens were p-type semiconductors. The variation of the mobility of the charge carriers with temperature has been derived and indicates that the scattering of the charge carriers is mainly due to thermal vibrations of the lattice except at low temperatures when impurity scattering becomes important. Reasonable agreement is obtained between the values of the mobility calculated from the change in conductivity in a magnetic field, and from Hall coefficient and conductivity measurements. The results of the thermoelectric power measurements, and the variation of the concentration of charge carriers derived from the Hall coefficient are discussed and compared with theory.

§ 1. INTRODUCTION

MOLYBDENITE is a semiconductor which is particularly suited for accurate measurements of electrical conductivity and Hall effect. It can be obtained in the form of large natural specimens which cleave into thin sheets. It is found that some of these specimens are homogeneous and appear to be single crystals. It is possible therefore to derive reliable information on such quantities as the concentration of charge carriers and their mobilities.

Previous investigators have found that the Hall effect of molybdenite is large and the sign negative, indicating that the charge carriers are electrons, although more recent work on other properties of molybdenite suggest that the carriers in some crystals can be 'holes'. Gottstein (1914) noted a large difference between the adiabatic and isothermal Hall effects, showing the presence of an Ettingshausen effect. Subsequently Heaps (1928) also observed a difference between the two Hall effects, but its magnitude was considerably smaller. It appears that there has been no investigation of the conductivity, in which allowance is made for contact resistance and inhomogeneity of the specimen. A change in resistance of molybdenite with magnetic field was noted by Heaps (1912), who found that the resistance decreases in the presence of a magnetic field. He mentions that much of the resistance lay in the contacts and admits that the change observed did not necessarily represent the behaviour of the substance in bulk. According to the theory of semiconductors the resistivity should increase with a magnetic field. Observations of the change in resistivity in a magnetic field are useful since they provide an independent method of calculating the mobility of the charge carriers and these results can be compared with those obtained from measurements of Hall effect and conductivity.

§ 2. EXPERIMENTAL PROCEDURE

Specimens were cut from samples of molybdenite having a rectangular shape except for narrow projections which could be used for potential probes and Hall probes.

The conductivity and Hall coefficient were measured by conventional methods over the temperature range 90°K to 800°K, and the variation of the thermoelectric power over the same temperature range was determined immediately afterwards.

The homogeneity of the specimens was tested by observing the potential distribution along a specimen when a constant current was passed through it.

For measurements of the variation of conductivity with magnetic field, a specimen was mounted on to a wax layer on one of the pole pieces of the magnet, the pole pieces in this case being tapered, and leads were attached to the specimen using Aquadag. With such an arrangement the separation between the pole pieces could be made very small and a field of 28 000 oersteds obtained. The usual method of determining a magnetic field with a search coil was unsatisfactory with such a small gap, and it was measured by Quincke's method, using a tube of distilled water between the pole pieces and observing the motion of the meniscus.

With this arrangement the transverse change in resistivity was measured, for the field at right angles to the basal plane. Measurements were also made in which the transverse change in resistivity was investigated with the magnetic field parallel to the basal plane.

§ 3. RESULTS

It is advisable, when performing measurements on natural crystals, to survey a large number of specimens and then select those most suitable for examination in more detail. Measurements were made of the conductivity and Hall effect on 23 specimens cut from crystals obtained from four different sources. The values of the Hall coefficient varied from 350 e.m.u. to 30 000 e.m.u. and the conductivity from 4·2 $\text{ohm}^{-1}\text{cm}^{-1}$ to 0·009 $\text{ohm}^{-1}\text{cm}^{-1}$. All the specimens except one were found to be p-type semiconductors. Five of these specimens were studied in detail. As will be shown later, specimens 5 and 21 appeared to be single crystals; specimens 14 and 15 had a very large Hall coefficient, indicating a small concentration of charge carriers; and specimen 22 was a large specimen to which good contacts could be made, resulting in a greater degree of accuracy for the measurement of the Hall effect. The variations of the conductivity and Hall coefficient with temperature are illustrated in figs. 1 and 2, and the potential distributions along specimens 5, 21 and 22 are shown in fig. 3.

Frequent checks were made to determine whether Ohm's law was obeyed by the specimens at different temperatures. Except for the upper end of the temperature range, where there was a sudden increase in conductivity, it was found that Ohm's law was obeyed for changes of several powers of ten of the potential difference across the specimen.

The transverse change in conductivity with the magnetic field at right angles to the basal plane for specimen 7 is illustrated in fig. 4. The Hall coefficient remained constant as the magnetic field was increased to 28 000 oersteds. The

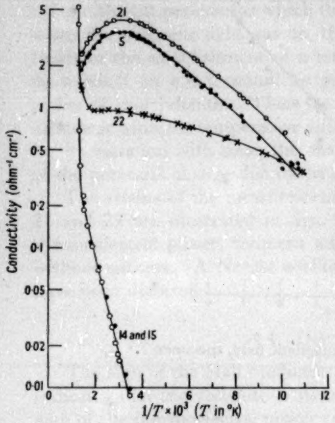


Fig. 1. Conductivity of specimens 5, 14, 15, 21 and 22.

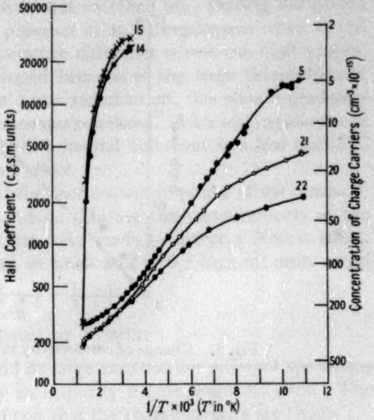


Fig. 2. Hall coefficient of specimens 5, 14, 15, 21 and 22.

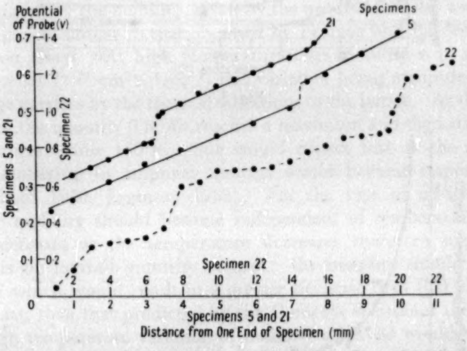


Fig. 3. Potential distribution along specimens 5, 21 and 22. Potential difference between a movable probe and one end of the specimen for a constant current flowing through the specimen.

change in conductivity of eight other specimens, including specimens 5, 21 and 22 was measured. The coefficient α defined by the equation

$$\Delta\sigma/\sigma=\alpha H^2 \qquad \ldots\ldots(1)$$

where $\Delta\sigma$ is the change in the conductivity σ due to a magnetic field H, was found to be similar for the different specimens. The mean value of α was

$2{\cdot}1 \times 10^{-12}$ oersted^{-2}. When the field was applied parallel to the basal plane, no change in resistivity could be detected. The value of α for this case was less than 4×10^{-13} oersted^{-2}.

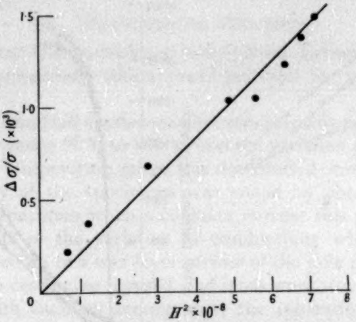


Fig. 4. Change of conductivity in a magnetic field, specimen 7.

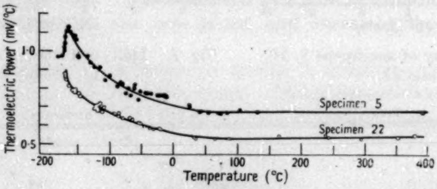


Fig. 5.

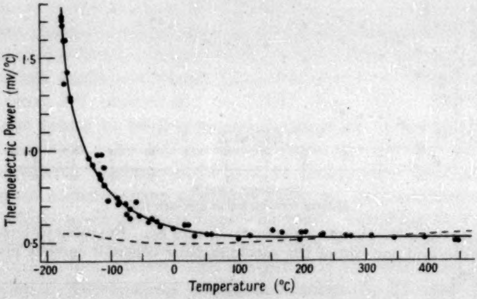


Fig. 6. Thermoelectric power of specimen 21. Broken curve—theoretical values.

The method used for measuring the Hall effect does not eliminate the error that might arise because of the presence of an Ettingshausen effect. Since previous investigators have observed a difference between the isothermal and adiabatic Hall effects, a careful attempt was made to investigate the magnitude of this error on specimen 21. A steady current was passed through the specimen and the potential difference across the Hall probes was balanced by a potentiometer. The variation of this potential difference with time was observed before and after a magnetic field was applied for 10 minutes. This procedure was repeated using different periods for which the field was switched on. During the period when the magnetic field was on, the presence of an Ettingshausen effect would result in the establishment of a temperature difference across the Hall probes. As a result an e.m.f. would be produced because of the large thermoelectric power of molybdenite. When the field was switched off, this should gradually disappear while the temperature difference was equalized. After making allowance for its variation with time, the change in potential difference was less than 1% of the potential change due to the Hall effect.

The results of the measurement of the thermoelectric power of specimens 5, 21 and 22 are illustrated in figs. 5 and 6. During the measurements of the thermoelectric power, frequent attempts were made to detect a Nernst effect, without success. A Nernst coefficient as small as 2×10^{-3} c.g.s.m. units could have been detected.

§ 4. DISCUSSION OF RESULTS

The sign of the Hall coefficient and its large variation for different specimens indicates that molybdenite is usually an impurity p-type semiconductor. The sign of the thermoelectric power confirms that the charge carriers are 'holes'.

(i) *Mobility of Charge Carriers*

The variation of the mobility, given by the product $0{\cdot}85R\sigma$, with temperature shown in fig. 7 is similar to that obtained by Pearson and Bardeen (1949) using silicon–boron alloys. At high temperatures the mobility v is proportional to $T^{-3/2}$ ($v = 8 \times 10^5 T^{-3/2}$ cm^{2}v^{-1}sec^{-1}) this variation being attributed to scattering of the charge carriers by the thermal vibrations of the lattice. As the temperature is decreased, the quantity $0{\cdot}85R\sigma$ reaches a maximum and then starts to decrease at temperatures below 150°K. One would expect that as the temperature is decreased scattering by impurity centres would become important (Conwell and Weisskopf 1950, Erginsoy 1950). For the case of un-ionized impurity centres the mobility should become independent of temperature and the fact that $R\sigma$ decreases as the temperature decreases therefore suggests that the scattering is by ionized impurity centres; the mobility should then be given by $R\sigma/1{\cdot}93$ which would result in a greater decrease than that shown in fig. 7, larger, in fact, than that predicted by the Conwell–Weisskopf theory.

The high temperature variation of mobility observed would be expected for metals and substances with an atomic lattice, but not for ionic substances for which one would expect an exponential law of the form predicted by Mott and Fröhlich (1939). However, in the case of molybdenite, which has a layer lattice, the measurements of the mobility have been confined to motion of the charge carriers parallel to the layers. Furthermore, Dutta (1944, 1945) has discussed the crystal structure of molybdenite and has pointed out that the interatomic distances observed along the basal plane do not support the presence of ionic bonding, and that there is evidence to suggest the presence of bonds of partially metallic character. It is not possible to correlate the numerical values of the mobility with theory because of insufficient information concerning the fundamental constants of molybdenite, and because of its highly anisotropic nature.

If the scattering of the charge carriers is due to thermal vibrations of the lattice their mobility should be the same for all specimens at a given temperature. The mobility calculated from $R\sigma$ of several specimens was considerably less than that found for specimens 5 and 21. There are two possible explanations for this discrepancy. Firstly, the graphs of the potential distribution along specimens 5, 21 and 22 show that specimen 5 is homogeneous except for a relatively small fault in the crystal. The resistance determined from the potential drop between the probes includes the resistance of this fault, and hence gives rise to an error in the calculation of the conductivity of the crystal. The same applies to specimen 21. The correction necessary for specimen 5 is 8% and for specimen 21 is 15%. For specimen 22 the graph shows that there are several faults in the crystal, and the error involved in neglecting them in the calculation of the conductivity will be large. This most probably explains the difference between the mobility of specimen 22 and that of specimens 5 and 21 shown in fig. 7. Secondly, for the pure specimens it is probable that the conduction is mixed, i.e. partly by holes and partly by electrons. In this case the mobility of the charge carriers is no longer given by the product $R\sigma$. For specimens 14 and 15 $R\sigma$ yields a value of the mobility of 22 $\mathrm{cm}^2\,\mathrm{volt}^{-1}\,\mathrm{sec}^{-1}$ at room temperature. As the temperature is increased the product $R\sigma$ increases until it is the same as for specimens 5 and 21 and then starts to decrease. This low value of $R\sigma$ may be due to the presence of mixed conduction at room temperature, and indicates that as the temperature rises the concentration of holes increases without a change in the concentration of electrons.

(ii) *Change of Conductivity in a Magnetic Field*

The decrease in conductivity in a magnetic field H provides us with an alternative method of calculating the mobility. According to Wilson (1936) the decrease should be given by the equation:

$$\Delta\sigma/\sigma = 0{\cdot}38v^2H^2 = \alpha H^2. \qquad \ldots\ldots(2)$$

The values of the mobility calculated from eqn. (2) for nine specimens are given in table 1 together with the values obtained from conductivity and Hall coefficient measurements.

Table 1. Mobility of Charge Carriers ($\mathrm{cm}^2\,\mathrm{v}^{-1}\,\mathrm{sec}^{-1}$)

Specimen No.	5	7	8	12	21	22	30	31	32
$1{\cdot}6\alpha^{1/2}$	204	228	212	222	206	233	338	216	259
$0{\cdot}85R\sigma$	144	24	53	5	147	31	63	23	45

The mean value of the mobility obtained from α was 237 ± 30 $\mathrm{cm}^2\,\mathrm{v}^{-1}\,\mathrm{sec}^{-1}$, and in view of the approximations involved, the agreement between this value and that obtained from the Hall coefficient and conductivity measurements of specimens 5 and 21 (162 $\mathrm{cm}^2\,\mathrm{v}^{-1}\mathrm{sec}^{-1}$) may be regarded as satisfactory.

The agreement of values of the mobility obtained for different specimens from magneto-resistance measurements would suggest that for semiconductors which are not available in the form of single crystals and for which, consequently, the values of the conductivity are uncertain, the change in conductivity in a magnetic field should provide a useful method for determining the mobility.

As no change in resistivity could be detected when the field was parallel to the basal plane, it would appear that the motion of the charge carriers is confined to the basal plane. A similar result was found by Dutta from measurements of the magnetic anisotropy of molybdenite at different temperatures.

(iii) *Concentration of Carriers*

Figure 2 shows the variation of the concentration n of the charge carriers with temperature, assuming that

$$R=3\pi/8n\boldsymbol{e}. \qquad \ldots\ldots(3)$$

If it is assumed that there is only one type of impurity centre, and that the activation energy ΔE of these centres is constant, then n should be given by

$$n^2=2{\cdot}41\times10^{15}T^{3/2}[\exp(-\Delta E/\boldsymbol{k}T)](n_{\mathrm{b}}-n). \qquad \ldots\ldots(4)$$

The variation of n with temperature predicted by this equation is compared in fig. 8 with the results obtained with specimen 5. Values of n_{b} and ΔE were chosen to give the best fit to the experimental curve and it is seen that the agreement between the theoretical and experimental curves is not satisfactory. For specimens 21 and 22 the discrepancy at low temperatures is greater. When it is remembered that at low temperatures the scattering of the charge carriers is probably due to ionized impurity centres, the concentration of charge carriers becomes $1{\cdot}93/R\boldsymbol{e}$. This would reduce the variation of n at low temperatures and increase the discrepancy.

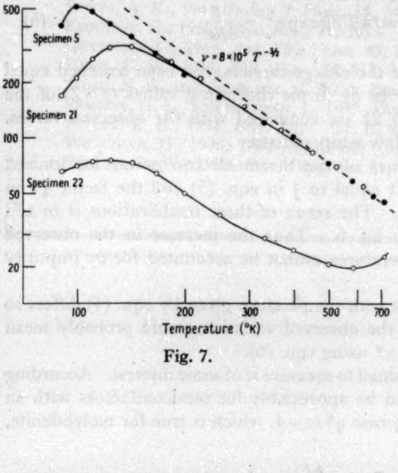


Fig. 7.

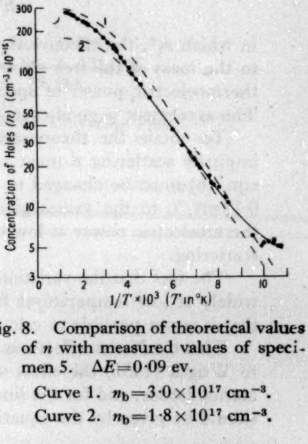


Fig. 8. Comparison of theoretical values of n with measured values of specimen 5. $\Delta E=0{\cdot}09$ ev.
Curve 1. $n_{\mathrm{b}}=3{\cdot}0\times10^{17}\ \mathrm{cm}^{-3}$.
Curve 2. $n_{\mathrm{b}}=1{\cdot}8\times10^{17}\ \mathrm{cm}^{-3}$.

At high temperatures, above 500°c, a very rapid increase in the conductivity was observed for the different specimens and in some cases this increase was permanent. The fact that the increase was permanent in some specimens suggests that the conduction at these temperatures is ionic rather than intrinsic. The rate of increase of conductivity was so high that it was not possible to take sufficient readings to determine the activation energy from the slope of the conductivity curve.

(iv) *Thermoelectric Power, Nernst Effect and Ettingshausen Effect*

The thermoelectric power dE/dT should be given (Wright 1951) by

$$\frac{dE}{dT} = -\frac{\boldsymbol{k}}{\boldsymbol{e}}(\beta - \eta^*) \qquad \ldots\ldots(5)$$

where η^* is the reduced chemical potential of the charge carriers and β depends on the variation of the mean free path l with the reduced energy η.

If $l = l_0\eta^r$ then

$$\beta = \frac{(2+r)F_{r+1}}{(1+r)F_r}, \qquad F_r = \int_0^\infty \frac{\eta^r\, d\eta}{1+\exp(\eta - \eta^*)}.$$

For a semiconductor with an atomic lattice, if lattice scattering predominates $r = 0$ and $\beta = 2$; if impurity scattering predominates, then for ionized impurity centres $r = 2$ and $\beta = 4$, and for neutral impurity centres $r = \frac{1}{2}$ and $\beta = 5/2$. Values of η^* have been determined from the Hall effect measurements using the formula

$$n = \frac{2}{\boldsymbol{h}^3}(2\pi m^* \boldsymbol{k} T)^{3/2} \exp \eta^* = \frac{3\pi}{8R\boldsymbol{e}} \qquad \ldots\ldots(6)$$

in which m^*, the effective mass of the charge carriers, has been assumed equal to the mass of the free electron. In fig. 6 the theoretical values ($\beta = 2$) of the thermoelectric power of specimen 21 are compared with the observed values. The agreement is good except at low temperatures.

To obtain the theoretical values of the thermoelectric power for ionized impurity scattering β must be put equal to 4 in eqn. (5) and the factor $\frac{3}{8}\pi$ in eqn. (6) must be changed to 1·93. The result of these modifications is to add 0·13 mv/°c to the values given in fig. 6. Thus the increase in the observed thermoelectric power at low temperatures cannot be accounted for by impurity scattering.

The fact that the variation of n with temperature given by eqn. (4) differs so widely at low temperatures from the observed variation would probably mean that it is not possible to calculate η^* using eqn. (6).

That the Nernst effect was too small to measure is of some interest. According to Wright (1951) the effect should be appreciable for semiconductors with an atomic lattice, and for the limiting case $\eta^* < -4$, which is true for molybdenite, should be given by the equation

$$Q = \frac{\pi}{4}\frac{\boldsymbol{k}l}{(2\pi m^* \boldsymbol{k} T)^{1/2}} \qquad \ldots\ldots(7)$$

where l is the mean free path of the charge carriers at the temperature T. When the value of l, calculated from the product $R\sigma$ assuming that $m^* = \boldsymbol{m}$ is substituted in this equation, it is found that the Nernst coefficient should be 8×10^{-3} c.g.s.m. at room temperature. An effect of this magnitude could have been readily detected.

A similar result is found for the Ettingshausen effect. When the formula given by Wright is used to calculate the Ettingshausen coefficient, and is combined with the measured value of the thermoelectric power, it is found that there should be a considerable difference between the adiabatic and isothermal Hall effects. However, Wright's analysis of the Ettingshausen effect is based on the assumption of thermal conduction by electrons, whereas in molybdenite the concentration of charge carriers is small and their contribution to the thermal conductivity will be insignificant compared with the lattice contribution (Putley 1952).

§ 5. CONCLUSION

The most useful result arising from the study of the electrical properties of molybdenite has been the evidence of the nature of the scattering of the charge carriers in a semiconducting binary compound. Over the temperature range for which lattice scattering is significant, the temperature variation of mobility agrees with that predicted by theory for a substance with an atomic lattice.

ACKNOWLEDGMENT

This work was aided by a grant from the research fund of the University of London.

REFERENCES

CONWELL, E., and WEISSKOPF, V. F., 1950, *Phys. Rev.*, **77**, 388.
DUTTA, A. K., 1944, *Indian J. Phys.*, **18**, 249; 1945, *Ibid.*, **19**, 225.
ERGINSOY, C., 1950, *Phys. Rev.*, **79**, 1013.
GOTTSTEIN, G., 1914, *Ann. Phys., Lpz.*, **43**, 1079.
HEAPS, C. W., 1912, *Phil. Mag.*, **24**, 813; 1928, *Ibid.*, **6**, 1283.
MOTT, N. F., and FRÖHLICH, H., 1939, *Proc. Roy. Soc.* A, **171**, 496.
PEARSON, G. L., and BARDEEN, J., 1949, *Phys. Rev.*, **75**, 865.
PUTLEY, E. H., 1952, *Proc. Phys. Soc.* B, **65**, 991.
WILSON, A. H., 1936, *Theory of Metals* (Cambridge : University Press).
WRIGHT, R. W., 1951, *Proc. Phys. Soc.* A, **64**, 984.

The Velocity and Temperature Dependence of Rubber Friction

By A. SCHALLAMACH
The British Rubber Producers' Research Association, Welwyn Garden City

MS. received 8*th December* 1952

Abstract. It is shown experimentally that the velocity of frictional gliding under constant tangential stress of rubber on glass and on silicon carbide cloth is, in the first approximation, an exponential function of the reciprocal absolute temperature and of the tangential stress. It is suggested that frictional gliding of rubber is a rate process.

§ 1. INTRODUCTION

THE coefficient of dynamic rubber friction has received little attention in the literature. Qualitative experiments by Ariano (1929, 1930) and Derieux (1934, 1935) showed that the coefficient of rubber friction increases with increasing sliding velocity, and this was confirmed by Roth, Driscoll and Holt (1942). These authors experimented with rubbers of the tyre tread type and found that the friction of such samples on plate glass increases slowly with increasing velocity, the rate of increase being lower than would be given by a logarithmic velocity dependence.

The published data on the temperature dependence of rubber friction are even scantier and appear to be the result of *ad hoc* experiments made in connection with specific technological problems. Breuil's (1910) experiments were inconclusive and Ariano (1929, 1930) found that the coefficient of rubber friction falls with increasing temperature.

We have investigated both velocity and temperature dependence of rubber friction by a method based on the fact that the positive velocity coefficient of rubber friction makes it possible to cause rubber to slide over a track at a constant velocity by application of a sufficiently high constant tangential stress. As mentioned above, the velocity coefficient is low so that relatively small variations of the tangential stress will produce large changes of the sliding velocity, and a wide range of velocities can be covered in this way without the necessity of elaborate gear boxes or similar devices. Further advantages of this method of studying rubber friction will become apparent later. This paper reports the results of measurements of the sliding velocity at various tangential stresses and temperatures.

§ 2. EXPERIMENTAL

The apparatus used for our experiments was a modified form of that described in an earlier publication (Schallamach 1952) to which reference is made for details of the shape of the samples. The geometrical contact area of the samples was in the present case 5·4 cm^2.

In order to control the temperature, the horizontal track was made hollow and water was pumped through it from a thermostat. The track was housed in a box lined with copper tubing which was connected in series with the hollow of the track so that an approximately uniform temperature was maintained within the enclosure. The temperature difference across the thickness of the sample did not exceed 0·25 centigrade degree.

The constant tangential stress was provided by lines which passed over free pulleys and carried hanging weights. The sliding velocity was automatically recorded by trailing behind the sample a paper strip on which a simple timing mechanism made marks at given intervals.

Preliminary experiments had shown that because of the sensitivity of the method special precautions had to be taken in order to obtain reproducible results. The first requirement was to reproduce the surface of the track as nearly as possible, and it was found that using ground plate glass and regrinding it after each run gave the most reliable results. The abrasive used for this purpose was carborundum, grade 1F. Additional measurements were made on silicon-carbide cloth 'John Bull' No. 150; by using fresh abrasive for each run reproducible results were obtained on this type of track.

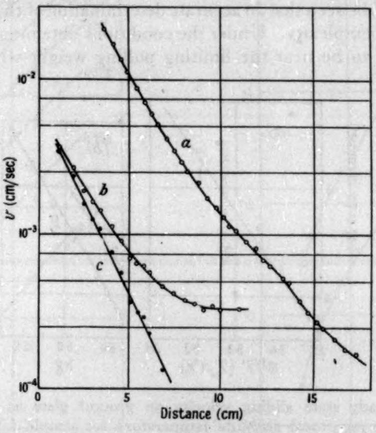


Fig. 1. Sliding velocity on ground glass plotted against distance travelled. Pulling weight: *a*, 5·8 kg, *b*, 7·8 kg. Curve *c*: curve *b* re-plotted as log [$(v-0{\cdot}00036)$ cm/s] against distance travelled.

In order to remove from the samples substances which might ooze out during the course of the experiment and produce random self-lubrication, the samples were extracted for 48 hours with alcohol-toluene azeotrope (Griffith *et al.* 1948), followed by 24 hours' extraction with alcohol, and subsequent vacuum drying. An unfortunate consequence of extraction of most types of rubber compound is that both natural and added antioxidants are removed thereby, and noticeable oxidation sets in within a matter of days. As it was not always possible to finish an investigation within this time, more than one sample had sometimes to be employed in the course of one investigation. All experiments were carried out with an unloaded vulcanizate of natural rubber of the following composition: rubber 100·0, ZnO 5·0, S 2·5, 'Santocure' (accelerator) 0·6, phenyl-β-naphthylamine (antioxidant) 1·0, stearic acid 1·0. The samples were cured for 50 minutes at a temperature of 142°c.

§ 3. RESULTS

(i) *Measurements on Ground Glass*

In our measurements, the normal load on the sample was always 6·44 kg. In view of the fact that the true tangential stresses depend on the actual area of contact between rubber and track (Schallamach 1952) the absolute value of which is not known, the results are given in terms of the pulling weights which actuated the samples.

When rubber slides on a surface which produces little abrasion, the sliding velocity decreases rapidly as the sample travels along the track. This is demonstrated for two different pulling weights of respectively 5·8 and 7·8 kg in fig. 1. In the case of curve *a* there is no indication of the velocity becoming constant, and eventually the sample did indeed stop. In case *b*, however, the sliding velocity became constant after a travel of about 10 cm. It is obvious that the pulling weight of 5·8 kg corresponds to less than the ultimate value of the coefficient of static friction, and it will be seen that an accurate determination of this quantity would be a matter of some complexity. Under the conditions obtaining in our experiments 7·8 kg was found to be near the limiting pulling weight which would produce steady motion.

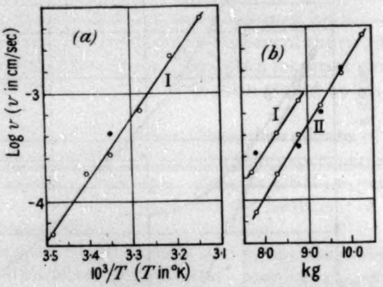


Fig. 2. Steady state sliding velocity on ground glass as a function of (*a*) the reciprocal absolute temperature for sample I with a pulling weight of 7·7 kg, (*b*) the pulling weight for samples I and II at 25°C.

When the pulling weight is increased in the course of the measurement the final value is attained very much sooner than when the measurements are begun with a new sample, and it appears that during sliding a conditioning of the surface is brought about. If the pulling weight is momentarily removed, the conditioning is partly lost. If, at the same time, the sample is lifted off the track, the conditioning virtually disappears. The practical importance of this effect is that much time is lost in re-conditioning the sample after it has traversed the whole length of the track and a new run is started. In view of the need for haste caused by the degradation of the samples by oxidation, a small alteration was made in the experimental arrangement. The track was divided into two equal sections in tandem, i.e. one section in front of the other, the butt joints being accurately ground square so as to present only a small discontinuity to the sample when travelling over it. After the sample had cleared the first section, this was removed for re-grinding; then the second section, with the sample on it, was pulled back and the first section inserted in front of it. This procedure made it possible to conduct the experiment without ever lifting the sample off the track, and to determine reproducible curves with one and the same sample.

The results showing the steady state velocities obtained with two samples are shown in fig. 2. The full circles in both graphs are repeat points. The curve of fig. 2 (*a*) is, within the limits of experimental error, a straight line; similarly, the graphs of fig. 2 (*b*) are very nearly linear apart from the region of the lowest pulling weights in the neighbourhood of the values corresponding to the static coefficient of friction.

Curves I and II in fig. 2 (*b*) illustrate the divergent absolute values of the sliding velocities which may be obtained with nominally similar samples; the slopes of the curves, on the other hand, appear to be very similar. The range of pulling weights within which steady sliding is possible had an upper limit of about 10 kg. If this limit was exceeded the sample accelerated slowly.

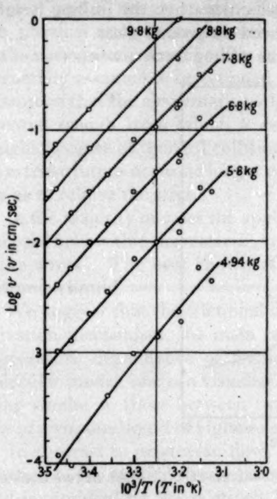


Fig. 3. Steady state sliding velocity on silicon-carbide cloth plotted against reciprocal absolute temperature for various pulling weights.

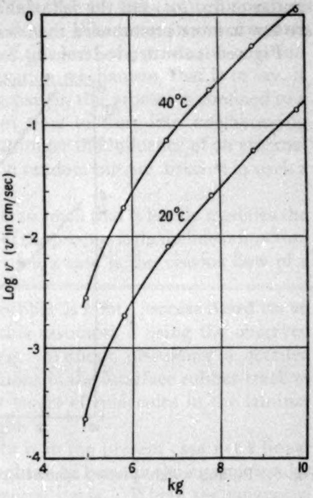


Fig. 4. Steady state sliding velocity on silicon-carbide cloth plotted against pulling weight.

Figure 2 shows that in the first approximation the sliding velocity of rubber on ground glass is (*a*) an exponential function of the reciprocal absolute temperature, (*b*) an exponential function of the pulling weight as long as this is sufficiently large compared with the weight necessary to overcome static friction. During the measurements, a certain amount of abrasion occurred and it was observed that mould marks on the surface of the sample were worn away; also, the sample left a faint trace on the track. The magnitude of the abrasion was, however, too small to be determined by weighing.

(ii) *Measurements on Silicon-Carbide Cloth*

Frictional sliding of rubber on silicon-carbide cloth depends far less critically on the experimental conditions than was found in the case of ground glass tracks.

In particular, there is no conditioning of the surface and the sample will slide at constant velocity almost immediately after it has been started. Also, steady running is possible in a wider region of pulling weights and velocities, this extension of the working range being mainly due to a lower static coefficient of friction. It might be pointed out here that the coefficient of friction is lower on the coarser surface of the abrasive than on the smooth surface of the ground glass because the local normal pressure where the rubber is indented by the particles of the abrasive is high (Schallamach 1952).

Figure 3 shows the results of the measurements of the sliding velocity on silicon-carbide cloth. The graphs have a slight curvature and can again in the first approximation be considered as linear. Figure 4 gives the logarithm of the sliding velocity at two constant temperatures as a function of the pulling weight, the points on this graph having been interpolated from fig. 3. This graph is similar to fig. 2 (*b*) but the fall of the sliding velocities when the pulling weights are low is more pronounced than was experienced on ground glass.

Figure 5 is also derived from fig. 3 and gives the pulling force as a function of the temperature at the velocities of 10^{-1} cm/sec and 10^{-2} cm/sec. The coefficient of friction is proportional to the pulling force and fig. 5 shows therefore the negative, but not constant, temperature coefficient of the coefficient of friction.

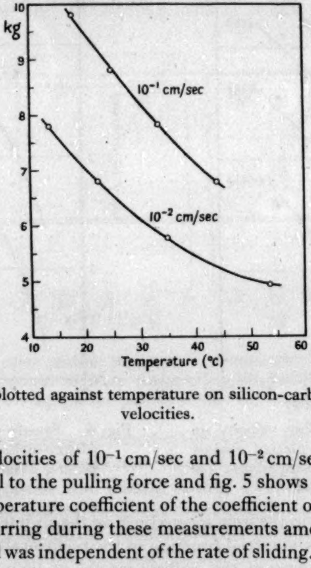


Fig. 5. Pulling weight plotted against temperature on silicon-carbide cloth at two constant velocities.

The abrasion occurring during these measurements amounted to about 17 mg for a run of 100 cm and was independent of the rate of sliding.

§ 4. DISCUSSION

The decrease of the sliding velocity of rubber on tracks which have only little abrasive effect (fig. 1) has its counterpart in the initial rise of the frictional force which is observed when friction measurements are made at constant velocity (Roth, Driscoll and Holt 1942). It has been suggested above that this effect is due to a conditioning of the rubber surface; its absence on an abrasive track where the surface is constantly renewed lends weight to this assumption. The effect arises most probably from an increase of the true area of contact between rubber and track under the combined influence of normal load, tangential stress and sliding. An increase of the true area of contact would lead to an enhanced frictional force as has been shown elsewhere (Schallamach 1952). A similar phenomenon in metallic friction has been demonstrated by McFarlane and Tabor (1950).

It may be observed that curve *b* of fig. 1 can be described by the simple mathematical relation $v_x - v = \text{const.} \exp(-x/l)$ where v is the final velocity, x is the distance travelled and l is a constant; this is shown by the linearity of curve *c* of fig. 1.

The most interesting feature of the results is, perhaps, the temperature dependence of the sliding velocity at constant pulling force F which can be approximately described by the equation:

$$v_{F=\mathrm{const}} = A' \exp(-E'/\boldsymbol{k}T) \qquad \ldots\ldots(1)$$

where A' and E' are constants. Equations of this type are familiar in connection with physical rate processes, such as viscous flow. Whenever an experimental result is adequately described by an equation of this kind, the process in question is generally assumed to be actuated by an activation mechanism, that is to say, it is assumed that the elementary particle taking part in the process is confined to a potential trough from which it escapes from time to time into neighbouring troughs because of thermal collisions, and that under the influence of an external stress these jumps occur no longer completely at random but are directed in such a way as to relieve the stress.

In the majority of cases the applied stress is so small that it hardly modifies the potential field within the material, and the rate of the process is then a linear function of the stress. The best known example of such a case is the viscous flow of a newtonian liquid.

We suggest that the frictional sliding of rubber is a rate process based on an activation mechanism, the main reason for this assumption being the observed temperature dependence of frictional sliding. Without discussing a detailed molecular model, one can visualize the conditions in the interface rubber-track as being similar to those between two adjacent layers of molecules in the laminar flow of a viscous liquid of high molecular weight.

In contrast to newtonian flow, the velocity is in the present case not a linear function of the applied stress, and this is thought to be because the intermolecular field is modified by the relatively large tangential stress. Where the tangential stress is sufficiently high compared with the value corresponding to the static coefficient of friction the experimental results can, in fact, be expressed by the following equation for v:

$$v = A \exp -(E - \gamma F)/\boldsymbol{k}T \qquad \ldots\ldots(2)$$

where A and γ are constants, F is the pulling force, and E is the activation energy measuring the height of the potential trough. It will be seen that on the basis of eqn. (2) the activation energy is not proportional to the slope of the curves of figs. 2 (*a*) and 3 which depends also on the tangential stress. By carrying out partial differentiation of eqn. (2) with respect to $1/T$ and F, the slope of the $\log v = f(1/T)$ curves is obtained as

$$[\partial \ln v/\partial(1/T)]_F = -E/\boldsymbol{k} + FT(\partial \ln v/\partial F)_T. \qquad \ldots\ldots(3)$$

The second term on the right-hand side of eqn. (3) should effect a decrease of the slope with increasing tangential stress. The accuracy of our experiment was not great enough definitely to confirm this prediction. The values of E calculated by means of eqn. (3) are given in the following table, $[\partial \ln v/\partial(1/T)]_F$ having been obtained from figs. 2 (*a*) and 3, and $(\partial \ln v/\partial F)_T$ from figs. 2 (*b*) and 4.

	Rubber on ground glass	Rubber on silicon carbide
Pulling weight (kg)	7·7	7·8
$-\boldsymbol{R}\ [\partial \ln v/\partial(1/T)]$ (cal/mole)	26600	20400
E (cal/mole)	19880	16160

The order of magnitude of the activation energies for frictional sliding given in the table is the same as that found for other activation processes in vulcanized natural rubber; for example, the activation energy for dielectric relaxation (Schallamach and Thirion 1949) is 17 kcal/mole. It is interesting to note that Eyring's (1936) theory of non-newtonian flow leads, for large tangential stresses, to an expression for the rate of flow which is formally identical with eqn. (2).

The explanation of our experimental results as it has been set out above, takes no cognizance of abrasion and in fact divorces abrasion completely from friction, in contrast to current theories of solid friction. We are not yet in the position to state definitely that frictional sliding of rubber can take place without any abrasion occurring although experiments with rubber on plate glass appear to suggest this possibility.

Another query arising out of our measurements on silicon-carbide cloth is what part of the frictional force is due to abrasion, and what part to friction proper, and how far the results are determined by the ratio of these two contributions. As the rate of abrasion is independent of the velocity, it is safe to say that the fraction of the tangential stress necessary to balance the abrasive effort is constant in the whole range of velocities employed by us, and that the corresponding correction of the graphs of fig. 4 consists in an horizontal displacement of the origin towards the right. The main points of our argument are not affected by this adjustment.

The absolute value of the contribution of abrasion to the measured frictional force is more difficult to estimate but a lower limit of this quantity can be given without difficulty. Assuming that each particle of the debris has been removed from the bulk rubber in one single action the total energy necessary to produce the debris can be calculated from the known energy density at break of the rubber. It is found in this way that the force necessary to produce the debris at the observed rate is 0·15 kg. It will be seen that, say, ten times this figure would still lead to a relatively small contribution to the total frictional force.

ACKNOWLEDGMENT

This work forms part of the research programme undertaken by the Board of the British Rubber Producers' Research Association.

REFERENCES

ARIANO, R., 1929, *Politecnico*, **10, 11**; 1930, *Rub. Chem. Tech.*, **3,** 287.

BREUIL, P., 1910, *Caoutchouc et Gutta-Percha*, **7,** 4257.

DERIEUX, B., 1934, *Elisha Mitchell Sci. Soc.*, **50,** 53; 1935, *Rub. Chem. Tech.*, **8,** 441.

EYRING, H., 1936, *J. Chem. Phys.*, **4,** 283.

GRIFFITH, T. R., STORY, E. B., BARKLEY, J. W. D., and MCGILVRAY, F. M., 1948, *Anal. Chem.*, **20,** 837.

MCFARLANE, J. S., and TABOR, D., 1950, *Proc. Roy. Soc.* A, **202,** 244.

ROTH, F. L., DRISCOLL, R. L., and HOLT, W. L., 1942, *J. Res. Nat. Bur. Stand., Wash.*, **28,** 439.

SCHALLAMACH, A., 1952, *Proc. Phys. Soc.* B, **65,** 657.

SCHALLAMACH, A., and THIRION, P., 1949, *Trans. Faraday Soc.*, **45,** 605.

Free Convection over Parallel Sources of Heat

By HUNTER ROUSE, W. D. BAINES AND H. W. HUMPHREYS
Iowa Institute of Hydraulic Research, State University of Iowa, Iowa City, U.S.A.

MS. received 23rd October 1952

Abstract. One phase of a fundamental investigation of convection over point and line sources of heat, conducted through the past decade at the Iowa Institute of Hydraulic Research, simulated a method of fog dispersal used for the landing of aircraft in England during the war. This involved the release of heat from parallel lines of burners, the convection from which produced characteristic patterns of mean flow and of heat intensity above the burners. Experiments described in this paper consisted of the measurement of the velocity distribution and the temperature distribution at various model scales and rates of heat generation. The results have been reduced to dimensionless diagrams of the two distribution functions which satisfy the elementary continuity and impulse-momentum requirements. From these diagrams generalized families of stream lines and isotherms for the mean motion above the sources have been plotted. The general functions reveal the basic pattern of convection and permit the approximate evaluation of velocity and temperature rise for any desired combination of the independent variables.

§ 1. INTRODUCTION

During the past war, parallel lines of petrol burners were used in England as a means of dissipating fogs over aircraft landing strips, and several test installations were also built in America. The basic principle involved was that of free convection resulting from the buoyancy of the heated air, the elevated temperature in the convection zone over the runway causing the fog in this vicinity to evaporate.

Because atmospheric conditions in the regions of prevalent fog are quite different in England and America, to be effective the burner installations had to differ accordingly. English fogs generally occur with essentially no wind, whereas those in America and in the Aleutians are often accompanied by winds of appreciable strength. As a result, the English systems consisted of equivalent lines of burners on both sides of the runway, and the American modifications involved a primary line some distance in the prevailing upwind direction with secondary lines on the downwind side.

The Iowa Institute of Hydraulic Research, as a part of its wartime programme, undertook at model scale the empirical evaluation of burner requirements for cross-wind conditions, and the project was reported upon in 1944 to the Office of Scientific Research and Development. The experimental data were later analysed as a fundamental problem in free convection (Rouse 1947). As the field appeared to be a fruitful one for further study, investigations were then made of the basic convection patterns over point and line sources without cross flow (Rouse, Yih and Humphreys 1952). One phase of the latter project dealt with the induced motion and the accompanying distribution of temperature over parallel line sources similar to those used in the English system of fog dispersal (Rankine 1950). Although the need for such a method may now have been eliminated entirely by the development of blind-landing techniques, the results of the investigation are presented herewith for purposes of record.

§ 2. PHYSICAL AND DIMENSIONAL CONSIDERATIONS

For conditions of two-dimensional convection over parallel sources of heat located at a horizontal boundary, the primary variables describing the mean flow are those indicated in fig. 1. Herein x and y are the vertical and horizontal coordinate positions with respect to a symmetrically located origin, y_0 is the distance from the origin to each of two sources of equal strength, u is the vertical and v is the lateral component of the mean velocity, and $-\Delta\rho$ is the local reduction in density corresponding to the temperature rise ΔT.

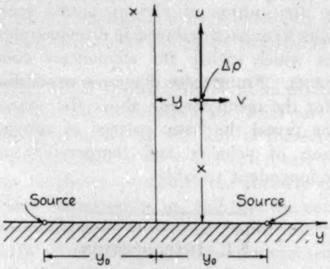


Fig. 1. Definition sketch.

If $\Delta\rho$ is assumed to be small in comparison with ρ, the density itself, the equation of continuity for constant-density conditions will apply with sufficiently close approximation:

$$\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}=0.$$

If it is further assumed that the pressure intensity is the same at all points, the equation of acceleration in the vertical direction will reduce to a form comparable with the Prandtl boundary-layer equation,

$$\rho u\frac{\partial u}{\partial x}+\rho v\frac{\partial u}{\partial y}=-g\Delta\rho+\frac{\partial \tau}{\partial y},$$

in which $-g\Delta\rho$ represents the buoyant force per unit volume which produces the acceleration and τ is the intensity of turbulent shear. Because v and τ are zero at the axis of symmetry and u, $\Delta\rho$, and τ vanish at a great distance from the axis, integration of the equation of acceleration over a horizontal plane in accordance with the equation of continuity will result in the following form of the customary momentum equation

$$\frac{d}{dx}\int_0^\infty \rho u^2\,dy=-\int_0^\infty g\Delta\rho\,dy. \qquad \ldots\ldots(1)$$

A similar equation can be written for the diffusion of heat on the assumption that the vertical rate of mixing due to turbulence is small in comparison with the lateral rate q:

$$u\frac{\partial(\Delta T)}{\partial x}+v\frac{\partial(\Delta T)}{\partial y}=-\frac{1}{c_p\rho}\frac{dq}{dy}.$$

Because q likewise vanishes both at the axis and at a great distance from it, integration of this equation over a horizontal plane will lead to the following condition of heat flux:

$$\frac{d}{dx}\int_0^\infty c_p\rho u\,\Delta T\,dy=0. \qquad \ldots\ldots(2)$$

The integral term itself, which is seen to be independent of elevation, thus represents the rate at which heat is generated per unit length of source.

The foregoing equations of momentum and heat flux not only specify conditions which must be approximately satisfied by the mean convection pattern, but they also define the variables which must be considered in an experimental investigation. The quantity $\Delta\rho$, first of all, can be related to the local temperature rise ΔT by the following simple expression for a perfect gas at constant pressure: $-\Delta\rho/\rho_0=\Delta T/T_0$. If this expression is introduced into eqn. (1) and the difference between ρ and the ambient value ρ_0 is again ignored, the momentum equation can be rewritten in the form

$$\frac{d}{dx}\int_0^\infty u^2\,dy=\int_0^\infty\frac{\Delta T}{T_0/g}\,dy. \qquad \ldots\ldots(3)$$

By defining the rate at which heat is produced per unit length of source as H/L, one can then express the vertical flux of heat in the corresponding form

$$\frac{H/L}{c_p\rho_0T_0/g}=\int_0^\infty u\frac{\Delta T}{T_0/g}\,dy. \qquad \ldots\ldots(4)$$

From the definition sketch of fig. 1 and the grouping of terms in eqns. (3) and (4), it will be seen that the simplest arrangements of the variables governing the velocity and temperature fields are as follows:

$$u=f_1\left(x,y,y_0,\frac{H/L}{c_p\rho_0T_0/g}\right)$$

$$\frac{\Delta T}{T_0/g}=f_2\left(x,y,y_0,\frac{H/L}{c_p\rho_0T_0/g}\right).$$

Inasmuch as only two-dimensional categories are represented by each series of five variables, the latter may be combined by means of the Π-theorem into functional relationships between three dimensionless parameters:

$$\frac{u}{A^{1/3}}=\phi_1\left(\frac{x}{y_0},\frac{y}{y_0}\right) \qquad \ldots\ldots(5)$$

$$\frac{\Delta T}{A^{2/3}T_0/gy_0}=\phi_2\left(\frac{x}{y_0},\frac{y}{y_0}\right) \qquad \ldots\ldots(6)$$

in which A is simply the group of terms with the dimension of (velocity)^3 appearing in the original arrangements of variables and indicating the strength of the source:

$$A=\frac{H/L}{c_p\rho_0T_0/g}.$$

These, then, are the relationships to be determined experimentally.

§ 3. Experimental Investigation

Measurements of the variables involved in eqns. (5) and (6) were carried out as an extension of a closely related laboratory investigation of the mean pattern of convection above a single line source of heat. Indeed, the latter problem represents the limiting case of the one now under discussion, as may be seen by writing the terms at the right of eqns. (5) and (6) in the alternative forms of y/x and y_0/x and assuming y_0/x to approach zero. Because of this fact, it was possible to use the same experimental equipment for both phases of the project.

Two-dimensional conditions free from external disturbance were obtained by confining the flow between vertical walls parallel to the xy plane. These walls were 8 feet long and 4 feet high and were placed either side of a very low platform 8 feet long and 4 feet wide. The top and the two end sections of the flow passage were open, and the unit was centred in a room 25 ft. by 15 ft. by 10 ft. in size. The heat source consisted of a specially fabricated gas burner 4 feet in length with a single line of $\frac{1}{16}$ inch holes in the trough-shaped top; the burner was set into the platform at mid-section with its upper edges flush. Commercial bottle gas was used as fuel; with the proper admixture of oxygen, a blue flame with minimum radiant energy was obtained. To simulate by the method of images two parallel sources of variable spacing, a 4 ft. by 4 ft. vertical partition was inserted between the parallel walls at any desired location of the plane of symmetry.

Measurements of the vertical component of the mean velocity were made by means of a sensitive vane anemometer $1\frac{1}{4}$ inches in diameter, the jewel bearings of which permitted velocities as low as 0·25 feet per second to be observed. The anemometer was calibrated in a 7-inch duct just beyond the rounded inlet, against velocity measurements made in the parallel jet from a nozzle several diameters downstream. Measurements of the temperature distribution were made with a copper–constantan thermocouple in conjunction with a potentiometer reading to 0·002 millivolt; the cold junction was maintained at room temperature outside the working section. The anemometer and the thermocouple could be mounted interchangeably on a traversing mechanism supported by rails along the upper edges of the two parallel walls.

Data were obtained by these means for various values of x, y, y_0 and burner output, the combination usually being that which would yield measurable values of u over a horizontal section of significant width. Apart from preliminary surveys to ensure that the flow was essentially two-dimensional, the measurements were made midway between the parallel walls. The magnitude of H/L for each run was evaluated from the observed distributions of u and ΔT by means of eqn. (4).

§ 4. Discussion of Results

All measurements of u and ΔT were generalized in accordance with the parameters of eqns. (5) and (6) and plotted to yield the desired functional relationships. The results of six typical runs representing two distinct geometric conditions are shown in fig. 2; herein the velocity data are plotted to the right and the temperature data to the left of the plane of symmetry. Although the individual points display the appreciable scatter which characterizes measurements of this nature, values for the same geometric conditions show no consistent variation with different rates of heat release. On the other hand the smooth curves approximating the trend of the data, and at the same time satisfying the requirements of eqns. (3) and (4) for momentum and heat flux, are seen to fall somewhat above the velocity data and below the temperature data in each case. This discrepancy was probably due in part to effects of radiation on the thermocouple indication and in part to the fact that the temperature differences were not always small (the average maximum reading was about 60°F above room temperature). However, previous studies had not disclosed a satisfactory method of shielding, and a detailed evaluation of thermodynamic effects was not considered compatible with the elementary nature of the project. For these reasons the results were accepted without further correction, and by correlation of the data from all runs the systematic sequences of curves satisfying eqns. (3) and (4) and shown in fig. 3 were obtained.

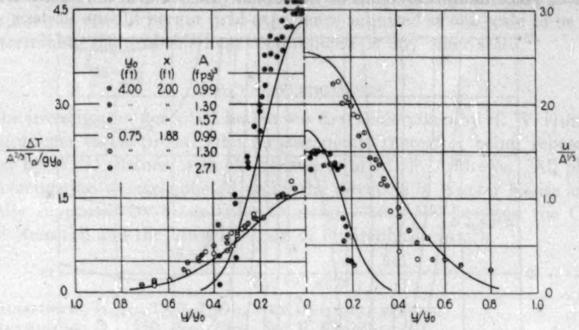


Fig. 2. Typical experimental results.

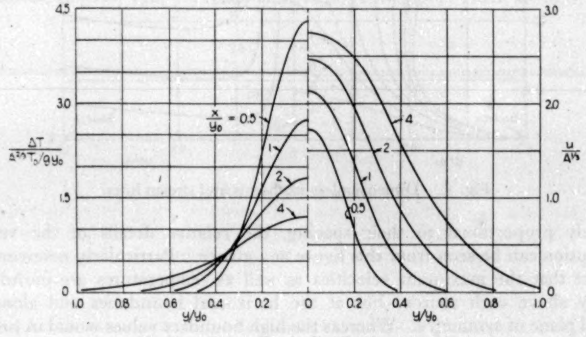


Fig. 3. Generalized distribution curves.

The curves for the relative distribution of temperature at various relative elevations at once permitted construction of the dimensionless pattern of isotherms reproduced in the left half of fig. 4. Herein, it should be noted, the region below $x/y_0 = 0{\cdot}5$ represents pure extrapolation, although the fact that all isotherms must begin at the source provides an approximate control over their form. The corresponding pattern of stream lines shown at the right of the figure was obtained from the systematized distribution curves of velocity by means of the definition equation for the stream function $\psi = \int u\,dy$. The pattern in the region below $x/y_0 = 0{\cdot}5$, as well as that to the right of the dip in each line, was again obtained through extrapolation. Inasmuch as the mean flow takes place in the direction of the stream lines and at a velocity which is

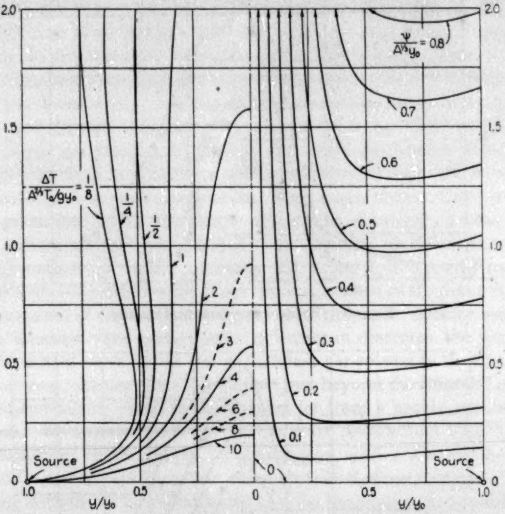


Fig. 4. Dimensionless isotherms and stream lines.

inversely proportional to their spacing, the relative details of the velocity distribution can be seen from this figure at a glance. Particularly noteworthy is the fact that the maximum velocities as well as temperatures are found—not directly above each source—but at the horizontal boundaries and along the vertical plane of symmetry. Whereas the high boundary values would in practice be reduced to some degree by boundary-layer deceleration, the fact remains that the zone of primary convection is in the immediate vicinity of the plane of symmetry.

The applicability of these results to the prediction of field conditions depends upon two limitations of the experiments themselves: their small scale, and the artificial means by which symmetry was obtained. The effect of scale is considered of less importance, for previous experiments with a point source had shown the Reynolds number of such motion to be far greater than that marking the transition from laminar to turbulent convection. Considerably more important is the fact that the degree of asymmetry which might be expected above two independent sources—even presuming them to be of identical strength—is in no way indicated. Indeed, slight winds in one direction or another (Rouse 1947) would probably offset any tendency towards symmetry which might otherwise exist.

Under the circumstances, the results presented must be regarded simply as a first approximation to the solution of a very complex problem. Surely, however, use of these functions to evaluate the order of magnitude of prototype conditions is preferable to the design of field installations with no generalized information at all. At the very least, the dimensionless parameters formulated in the analysis should permit field experience acquired at one scale to be utilized in determining the probable heat requirements at any other scale.

ACKNOWLEDGMENTS

The investigation described herein was first undertaken by H. W. Humphreys as a graduate thesis project, the measurements thereafter being repeated and refined by W. D. Baines, with the assistance of M. F. Andrews. All phases of the investigation were conducted under the direction of Hunter Rouse and were partially supported by funds from Contract N8onr-500 between the Office of Naval Research and the Iowa Institute of Hydraulic Research.

REFERENCES

HUMPHREYS, H. W., 1950, *Thesis*, State University of Iowa.
RANKINE, A. O., 1950, *Proc. Phys. Soc.* B, **63,** 225.
ROUSE, HUNTER, 1947, *J. Appl. Mech.*, **14,** No. 3, A-225.
ROUSE, HUNTER, YIH, C. S., and HUMPHEYS, H. W., 1952, *Tellus*, **4, No.** 3, 201.

Sorption of Gases at Very Low Pressures by Thorium Powder

By S. WAGENER

Electronics Division, Post Office Research Station, London N.W.2

MS. received 8th October 1952

Abstract. The rate of sorption of thorium powder for oxygen and hydrogen is measured at pressures between 10^{-7} and 10^{-3} mm mercury as a function of temperature of the thorium and of the period of exposure to gas. The sorbed quantities of gas are determined from the variation of rate of sorption with time. It is found that oxygen is sorbed irreversibly, forming a stable oxide with a dissociation pressure which is immeasurably low. The activation energy of this chemisorption is ascertained from the increase of rate of sorption with temperature of the thorium and is found to be about 0·75 kcal/mol.

Hydrogen is sorbed reversibly by a process of solution. The correlations between sorbed quantity of hydrogen, hydrogen pressure and temperature of the thorium are those to be expected for an equilibrium of solution. During dissolution of hydrogen heat is evolved and a value of about 13 kcal is found for this heat of solution. There are indications that at very low pressures most of the hydrogen in a tube containing thorium powder is in the atomic state. The rate of sorption for hydrogen has a maximum value between 700 and 750°K.

The significance of these results for the application of getter materials in electronic valves is discussed briefly.

§ 1. INTRODUCTION

During recent years thorium metal powder has been used with some measure of success for reducing the residual gas pressure in electronic valves. The thorium powder was deposited as a thin coating on one of the electrodes of the valves concerned. After an initial outgassing process the powder was found capable of efficiently sorbing deleterious gases at pressures of 10^{-7} mm Hg and lower. The physico-chemical processes on which this sorption is based have been investigated in some detail and an account of these investigations is given below.

§ 2. METHODS OF MEASUREMENT

Sorption of gases is normally investigated by directly measuring the sorbed quantities of gas using either a volumetric or a gravimetric method. Such methods, however, fail at the very low pressures concerned since the quantities to be measured become too small to be detected in this way. The investigations to be described here were based on a method of measuring rates of sorption which has been described elsewhere (Wagener 1950). The principle of this method is illustrated in fig. 1.

Gas is injected into the vessel containing the thorium via a capillary and the pressures p_p and p_t on either side of this capillary are measured by two ionization gauges IG1 and IG2. Then, from the values p_p and p_t and from the flow resistance F of the capillary, the rate of sorption is obtained as

$$S=\frac{p_p-p_t}{p_t}F. \qquad \ldots\ldots(1)$$

If the rate of sorption is measured as a function of time t, the sorbed quantity Q_s can be obtained using the formulae

$$Q_s=\int_0^t p_t\, S\, dt \qquad \ldots\ldots(2)$$

or

$$Q_s=F\int_0^t (p_p-p_t)\, dt. \qquad \ldots\ldots(3)$$

Two types of capillaries were used in the experiments. One of these, the standard one, had 2 mm bore, and its conductance for oxygen as measured in a previous calibration (Wagener 1950) was $F=17{\cdot}5$ cm^3/second. The second capillary which had a smaller bore was used for hydrogen in order to increase the pressure gradient along the capillary. Its conductance was 5·7 cm^3/sec for oxygen and 23 cm^3/sec for hydrogen.

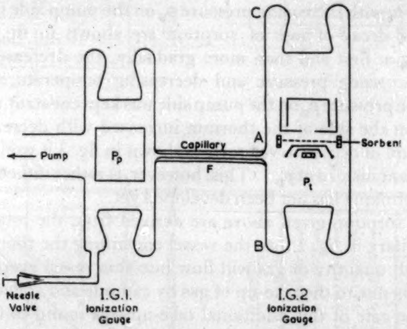


Fig. 1. Experimental arrangement for measurement of rates of sorption.

The thorium powder employed had a particle size between 1 and 10 μ. It was deposited electrophoretically by a method described elsewhere (Espe 1948) on pre-outgassed nickel sleeves of the type normally used for oxide cathodes. The method of depositing gave a very smooth thorium coating of normally 5 mg weight on a sleeve area of 2 cm^2. After deposition the coated sleeves were outgassed and slightly sintered in a high vacuum by heating up to 1050°K for 90 seconds. Two such sleeves, after being supplied with alumina-coated tungsten heaters, were located in the ionization gauge IG2 on the far side of the capillary in a manner to be seen from fig. 1.

The measurements of sorption were undertaken on a diffusion pump providing a pressure of 10^{-7} mm mercury after baking the glassware for 1 hour at 400°C. Subsequent to this bake all metal components of the two ionization gauges were outgassed at about 900°C either by eddy current heating or by electron bombardment. The getters were outgassed at the same temperature previous to every series of measurements (Wagener 1952).

§ 3. RATE OF SORPTION FOR OXYGEN

When determining the rate of sorption, the pressure p_p on the pump side of the capillary was increased gradually by admitting gas and the pressure p_t at the thorium was measured as a function of p_p. Then, for oxygen, a curve as shown by the broken line in fig. 2 was obtained. This curve could be transformed into a straight line by deducting the initial pressure on the getter side, obtained without admitting oxygen from outside (point on the extreme left of fig. 2) from all the appropriate pressures measured subsequently (plus crosses in fig. 2).

The straight lines plotted in fig. 2 are significant because they show that the rate of sorption of oxygen as derived from formula (1) is nearly independent of pressure. The transformation leading to these straight lines can be explained physically by assuming that a residual pressure of a gas which is not oxygen exists above the thorium. The nature of this gas, which remains unaffected during the entire measurement, is discussed in § 7.

The straight lines of fig. 2 are only obtained if the measuring period is kept short. If longer periods are used, the rate of sorption decays, in particular at the higher pressures. Owing to this decay the lines shown in fig. 2 will bend off towards higher pressures p_t with increasing pressure p_p on the pump side (Wagener 1952).

Details of the decay of rate of sorption are shown in fig. 3. The rate S decreases rapidly at first and then more gradually, the decrease becoming more marked with increasing pressure and decreasing temperature. During these measurements the pressure p_p on the pump side was kept constant and consequently the pressure p_t on the side of the thorium increased with decreasing S. For an accurate evaluation of decay curves such as shown in fig. 3 it would be desirable to maintain p_t constant instead of p_p. This, however, is rather difficult experimentally and a suitable technique has not been developed yet.

The rates of sorption given above are derived from the total rate of gas flow through the capillary in fig. 1 into the vessel containing the thorium. It may be noted that a small quantity of gas will flow into this vessel even if it contains no thorium. This is due to the take-up of gas by cathode and collector of the ionization gauge. The rate of this additional take-up was found to be approximately 30 cm^3/sec for oxygen and 5–10 cm^3/sec for hydrogen.

§ 4. RATE OF SORPTION FOR HYDROGEN

The rate of sorption S of thorium for hydrogen is smaller than that for oxygen. Typical curves, showing S_{H_2} as a function of time for different pressures p_p on the pump side, are given in fig. 4. When comparing figs. 4 and 3, it will be seen that the sorption rate decays much more quickly for hydrogen than it does for oxygen. Owing to this quick decay it has not been possible to obtain accurate values of the initial rate of sorption at time $t = 0$.

The rate of sorption for hydrogen increases with temperature up to about 750°K where a maximum is reached. Beyond this temperature a decrease in S with increasing value of T is observed (see figs. 5 and 6). This behaviour is very different from that which was found for oxygen (see also § 6).

Curves as shown in figs. 4, 5 and 6 were obtained with several samples of thorium although the absolute values of S were lower in some cases than those given in the figures.

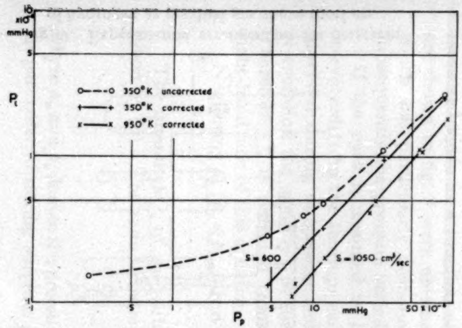


Fig. 2. Pressure p_t above thorium as a function of pump pressure p_p.

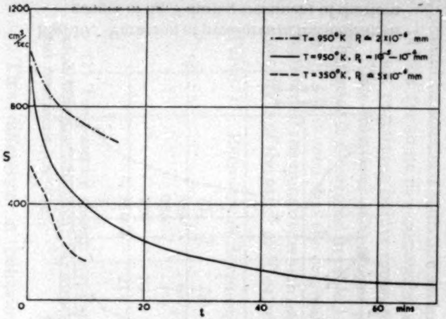


Fig. 3. Decay of sorption rate of oxygen.

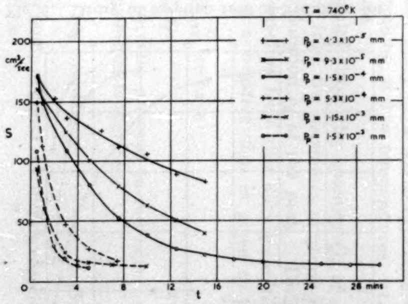


Fig. 4. Decay of sorption rate of hydrogen for different pressures.

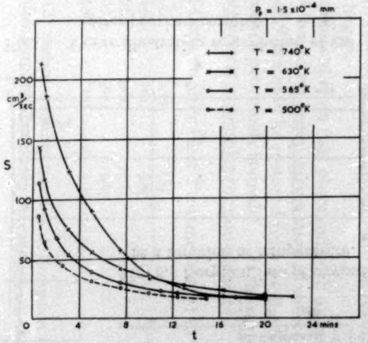


Fig. 5. Decay of sorption rate of hydrogen for different temperatures of thorium (T=500–740°K).

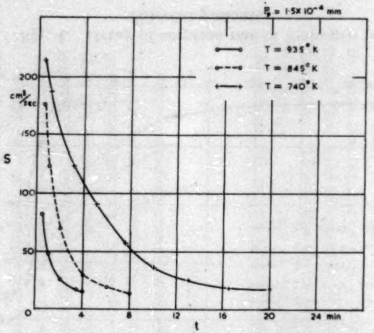


Fig. 6. Decay of sorption rate of hydrogen for different temperatures of thorium (T=740–935°K).

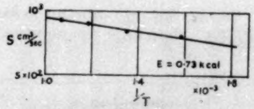


Fig. 8. Sorption rate of oxygen as a function of temperature.

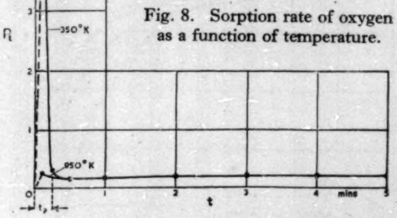


Fig. 7. Curve illustrating re-evolution of gas after previous sorption.

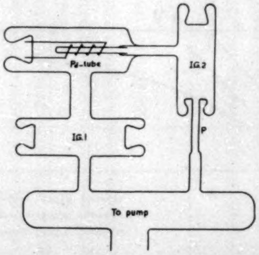


Fig. 9. Experimental arrangement for detection of hydrogen as residual gas above thorium.

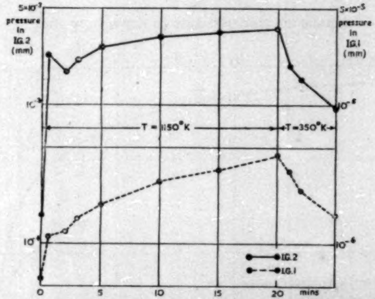


Fig. 10. Variation of pressures in the ionization gauges of fig. 9 during treatment of thorium at different temperatures.

§ 5. REVERSIBILITY OF SORPTION

In the case of hydrogen a series of decay curves such as those of figs. 4, 5 and 6 could be obtained with one sample of thorium since heating the thorium to about 1150°K for some five minutes enabled the condition of the thorium to be restored to that pertaining before measurements began.

This behaviour of hydrogen is again different from that observed with oxygen. The latter is illustrated in table 1 which is based on decay curves such as shown in fig. 3. It is seen that the original value of S can only be restored fully by the heat treatment if the quantity which has been sorbed previously is small. If the sorbed quantity is of the order of 10 or greater recovery is incomplete or non-existent.

Table 1. Recovery of Sorption Rate of Oxygen after Heat Treatment at 1150°K.

(1)	(2)	(3)	(4)	(5)
350	0·7	560	100	580
950	1·1	1015	660	1030
950	17	1030	280	650
950	21	650	100	100

(1) Temperature during sorption (°K); (2) sorbed quantity ($\mu \times$ l.); (3) sorption rate before sorption (cm³/sec); (4) sorption rate after sorption (cm³/sec); (5) sorption rate after heat treatment (cm³/sec).

This difference in behaviour between hydrogen and oxygen led to an investigation of reversibility of the sorption process. For this the thorium was exposed to gas at a certain pressure p_t for a time t_s. Subsequently the temperature of the thorium was raised to 1150°K and the pressure p_t was measured again as a function of time. In this way outgassing curves such as are shown in fig. 7 were obtained. An initial sharp peak was always observed which was particularly pronounced after sorbing at low temperatures, whilst subsequently the pressure varied only slightly.

If the peak is assumed to be of triangular shape the quantity Q_r of gas re-evolved from the thorium can be derived from the maximum pressure $p_{\max}$ observed during outgassing and from the duration t_p of the peak. During re-evolution the gas flows in the opposite direction from that during sorption. Therefore Q_r is obtained from eqn. (3) if a negative sign is given to the right-hand side of this equation.

Neglecting p_p ($p_p \ll p_{\max}$) and with $p_t = \frac{1}{2} p_{\max}$, we have

$$Q_r = \tfrac{1}{2} F p_{\max} t_p. \qquad \ldots\ldots (4)$$

The quantity Q_s of gas sorbed by the thorium previous to re-evolution is given by (3). The ratio between the two quantities is therefore

$$\frac{Q_r}{Q_s} = \tfrac{1}{2} p_{\max} t_p \Big/ \int_0^{t_s} (p_p - p_t)\, dt \qquad \ldots\ldots (5)$$

where p_p and p_t denote the pressures on either pump or getter side during sorption. In particular, if during sorption $p_t \ll p_p$ we have

$$\frac{Q_r}{Q_s} \simeq \frac{1}{2} \frac{t_p}{t_s} \frac{p_{\max}}{p_p}. \qquad \ldots\ldots (5a)$$

Values of the 'reversibility factor' Q_r/Q_s which have been measured so far are given in table 2. Values $Q_r/Q_s = 1$, showing that the sorption is completely reversible, are only obtained for hydrogen. For oxygen the appropriate values are smaller than 0·01 indicating that the preponderant part of this gas is sorbed irreversibly and cannot be removed at temperatures up to 1150°K.

Table 2. Reversibility Factors of Oxygen and Hydrogen

Gas	Temp. of Th (°K)	p_t (10^{-8} mm)	t_s (min)	Q_r/Q_s
Oxygen	350	2·5	3	0·005
	350	4	10	0·004
	350	4	15	0·006
	950	2	10	0·0004
	950	2	15	0·0003
	950	20	30	0·0005
Hydrogen	350	8	5	1·0
	350	8	5	1·0

After sorption of oxygen and heating at 1150°K the pressure above the thorium returned to about the same low value which had existed before sorption although no appreciable quantity of sorbed oxygen had been removed. This shows that, in equilibrium and within the limits of measurability, the oxygen pressure above thorium plus sorbed oxygen is the same as above thorium containing no oxygen. It is concluded from these measurements that the system thorium–oxygen represents a case of chemisorption in which a stable oxide, probably thorium dioxide (ThO_2), is formed which has a very low dissociation pressure.

The system thorium–hydrogen, however, may be considered as an equilibrium of solution in which hydrogen can be injected into thorium and removed at will. The experiments described below will give further proof that these assumptions are justified.

§ 6. TEMPERATURE DEPENDENCE OF SORPTION OF OXYGEN (MEASUREMENT OF ACTIVATION ENERGY)

Further evidence on the mechanism of the sorption of oxygen can be obtained from its temperature dependence. If this sorption is due to a chemical process one would expect its rate to vary with temperature according to an exponential law

$$S = B \exp(-E/\boldsymbol{R}T) \qquad \ldots\ldots(6)$$

where B is a constant, E is the activation energy and $\boldsymbol{R}$ the universal gas constant. Log S, therefore, should be a linear function of $1/T$ which is verified by fig. 8. The activation energies E obtained with three different samples of thorium are given in table 3.

Table 3. Activation Energies E of Different Samples of Thorium

Sample No.	1	2	3
E (kcal)	0·63	0·9	0·73

The comparatively low value of the activation energy may be due to the fact that the gas molecules are pre-activated under the impact of the electrons in the ionization gauges. Such an activation, presumably consisting of dissociation of molecules into atoms, was observed when investigating barium as a sorbing material (Wagener 1951). It is hoped to avoid such activation by using a Knudsen gauge when it is available.

§7. THE RESIDUAL HYDROGEN PRESSURE ABOVE THORIUM

It is known from the investigations of Sieverts and Roell (1926) at pressures between 3 and 760 mm that considerable quantities of hydrogen can be dissolved in thorium. The dissolved quantities are correlated thermodynamically with the hydrogen pressure outside the thorium. Small quantities of hydrogen, therefore, which have not been removed during previous heat treatment should produce a residual hydrogen pressure. In view of this, it may be expected that the residual pressure which was observed when measuring the sorption rate of oxygen and which produced the curvature of the lines representing S plotted against pressure p (see fig. 2), is due to hydrogen.

It has been attempted to prove this conclusion by using the experimental arrangement shown in fig. 9. The sleeves coated with thorium were located inside an ionization gauge IG2 in the usual way. During the outgassing phase this ionization gauge was pumped directly through the pumping tube P. When outgassing was completed P was sealed off and the only remaining connection to the pump was through a palladium tube and via another ionization gauge IG1.

After these preparations the palladium tube was heated by a surrounding tungsten spiral in order to enable hydrogen to pass through. The temperature of the thorium was then raised from its value in the unheated state (~350°K) to 1150°K and the pressure in the two gauges measured as a function of time. Two typical curves obtained are shown in fig. 10. It is clearly seen that the rise in pressure above the thorium (in IG2) which is due to the increase in temperature, produces a rise in pressure in the gauge on the other side of the palladium tube (IG1) Conversely, when with reduced temperature of the thorium the pressure above the thorium falls, the pressure on the far side of the palladium tube falls as well. This experiment clearly shows that the residual pressure above the thorium is due to a gas which can easily pass through palladium. This gas can only be hydrogen which most likely originates from the processes used for preparing the thorium.

The general assumption is that hydrogen is dissolved in metals like thorium as atoms (see Fowler and Smithells 1937). If therefore the dissolved hydrogen is re-evolved into the vacuum, the hydrogen will appear there in atomic form and will only recombine to molecules after a certain recombination time has elapsed. This recombination time will increase with decreasing number of collisions between hydrogen atoms and hence with increasing mean free path. Since the mean free path at the low pressures concerned is very long, it appears likely that a considerable part of the hydrogen existing above the thorium consists of atoms.

Langmuir (1950) has shown that atomic hydrogen can be detected by its reducing effect on tungsten oxide. In order to make use of this method of detection, two groups of four nickel sleeves, one group coated with thorium and the other one not, were sealed into glass bulbs on whose surface a thin film of a blue tungsten oxide, probably W_2O_5, had been deposited. The sleeves were run in these bulbs at a temperature of 1050°K in a hydrogen atmosphere of 5×10^{-3} mm pressure for about 60 hours. After this period the film on the bulb containing the uncoated sleeves had retained its original blue colour while the colour of the film on the second bulb, containing the sleeves coated with thorium, had definitely turned black. This was seen as an indication that the tungsten oxide on the second bulb had been reduced to metallic tungsten by hydrogen atoms which had been formed during dissolution in the thorium.

§ 8. THE THORIUM–HYDROGEN EQUILIBRIUM

Let us consider the equilibrium of solution of a gas which exists as atoms both inside and outside a solvent metal. Let q_s denote the quantity of gas dissolved per unit weight of metal ($q_s = Q_s/w$, where w is the weight of metal). The correlation between quantity dissolved and gas pressure p_{at} then is (Ulich 1948)

$$q_s/p_{at} = K_s \qquad \ldots\ldots(7)$$

where K_s is the appropriate equilibrium constant.

Using the well-known equation (Glasstone 1940)

$$\frac{\partial \ln K_s}{\partial T} = \frac{H_{at}}{\boldsymbol{R}T^2} \qquad \ldots\ldots(8)$$

in which H_{at} denotes the heat of solution per gram-atom of gas, we have:

$$\frac{\partial \ln (q_s/p_{at})}{\partial T} = \frac{H_{at}}{\boldsymbol{R}T^2} \qquad \ldots\ldots(9)$$

or

$$\ln (q_s/p_{at}) = Z - \frac{H_{at}}{\boldsymbol{R}T}. \qquad \ldots\ldots(10)$$

In this formula a negative value of H_{at} indicates that heat is evolved during dissolution.

If, contrary to the assumption above, the gas outside the metal consists of molecules, the equilibrium between molecules and atoms has to be considered as well. This is given by:

$$\frac{{p_{at}}^2}{p_{mol}} = K_D, \qquad \ldots\ldots(11)$$

K_D denoting the partial pressure of the molecules. Introducing p_{at} in (10) and denoting the heat of solution per gram-mole by $H_{mol} = 2H_{at}$, we obtain

$$\ln \{q_s/(p_{mol})^{1/2}\} = Z_{mol} - H_{mol}/2\boldsymbol{R}T. \qquad \ldots\ldots(12)$$

If two special cases, that of constant temperature and that of constant quantity dissolved, are considered, we have from either (10) or (12):

(*a*) $T = const.$

for atoms: $p = aq_s$ (Henry's law) ......(13 *a*)

and for molecules: $p = a^2 {q_s}^2$. ......(13 *b*)

Such a q^2-law was for instance confirmed for the solution of hydrogen in titanium by Kirschfeld and Sieverts (1929) at pressures between 8 and 760 mm (for $T = 1000°\text{c}$).

(*b*) $q_s = const.$

for atoms: $p = k \exp (H_{at}/\boldsymbol{R}T)$ ......(14 *a*)

and for molecules: $p = k^2 \exp (H_{mol}/\boldsymbol{R}T)$. ......(14 *b*)

In order to check eqns. (13) the equilibrium pressure p has to be measured as a function of the dissolved quantity Q_s. For this purpose measurements of the rate of sorption, such as represented by figs. 4, 5 and 6, can be utilized. The sorbed quantity Q_s can be ascertained from the curves by an integration using eqn. (3) which by combination with (1) can be transformed into

$$Q_s = p_D F \int \frac{S}{S+F} dt. \qquad \ldots\ldots(15)$$

There is, however, one difficulty. In order to obtain the equilibrium pressure appropriate to Q_s from the normal pressure measurements the decay curves $S=f(t)$ have to be extended until the rate of sorption S is zero. Only then is equilibrium established and the equilibrium pressure

$$p_{\mathrm{equil}}=p_{\mathrm{p}}=p_{\mathrm{t}}. \qquad \ldots\ldots(16)$$

Since the decay curves (see fig. 4 etc.) tail out very slowly, such a measurement would take a very long time during which the sorption rate S could only be measured very inaccurately on account of its smallness (see §9). Hence determination of the sorbed quantity Q_s by using eqn. (15) would also be very inaccurate.

In order to avoid this difficulty S was only measured until a value of between 10 and 20 cm³/sec was obtained. Subsequently the needle valve was closed gradually and the pressure on the pump side was thus reduced until it equalled the pressure on the side of the thorium and finally became smaller than the latter.

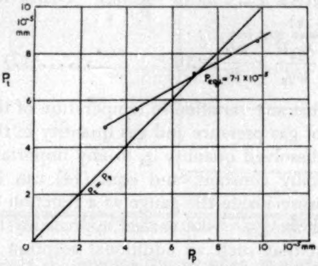


Fig. 11. Pressure p_{t} above thorium as a function of pump pressure p_{p} during determination of hydrogen equilibrium pressure.

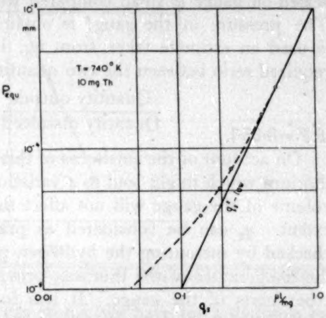


Fig. 12. Equilibrium pressure of hydrogen above thorium as a function of quantity of hydrogen dissolved.

The equilibrium pressure p_{equil} could then be ascertained by a graphical interpolation method. Figure 11 in which p_{t} is plotted as a function of p_{p} for such a run of measurements, illustrates the procedure. The curve representing $p_{\mathrm{t}}=f(p_{\mathrm{p}})$ is intersected by the line $p_{\mathrm{t}}=p_{\mathrm{p}}$ and the intersection point gives the value of the equilibrium pressure according to (16). This equilibrium pressure is not exactly appropriate to the value of Q_s which is derived from curves like those in fig. 4, because, until $p_{\mathrm{t}}=p_{\mathrm{p}}$, a small additional quantity of hydrogen is sorbed by the thorium. The time necessary to establish $p_{\mathrm{t}}=p_{\mathrm{p}}$ in this case, however, is much smaller than by continuing the measurement of the decay curve, and hence the accuracy obtained is correspondingly greater. The method was found satisfactory at least for determining roughly the correlation between equilibrium pressure and sorbed quantity.

Figure 12 shows the results obtained. A line representing the q^2-law of eqn. (13 *b*) has been plotted in the graph. It is seen that with decreasing pressure the measured points deviate more and more from this line and approach the q^1-line. This can be attributed to an increasing number of atoms in the hydrogen above the thorium owing to increasing mean free path and decreasing recombination. It may be expected, therefore, that at pressures below the range within which measurements could so far be performed, the hydrogen above the thorium will consist of atoms only.

The existence of a correlation between pressure and sorbed quantity confirms the concept of an equilibrium of solution between thorium and hydrogen. There is, at the pressures concerned, no indication of formation of a hydride which would have a dissociation pressure independent of the quantity of hydride formed.

In order to confirm that eqns. (14) also hold, a different experimental arrangement was used. Ionization gauges similar to those shown on the right-hand side of fig. 1 (IG2) were employed. After the normal pumping process a small but unknown quantity of hydrogen was dissolved and retained in the thorium; the gauge was subsequently sealed off the pump. It can be estimated that in such a case the quantity of hydrogen in the volume V of the sealed-off gauge is small compared with the quantity dissolved in the thorium. The pressure in the gauge is obtained from eqn. (13 *a*) or (13 *b*). If (13 *a*) is used an estimate taken from fig. 12 gives $a \simeq 0{\cdot}2\mu/(\mu \times \text{l./mg})$. Hence the required ratio between the two quantities is:

$$\frac{\text{Quantity outside}}{\text{Quantity dissolved}} = \frac{a q_s V}{w q_s} = \frac{aV}{w} \simeq 10^{-3} \qquad \ldots\ldots(17)$$

if $V = 0{\cdot}05$ l.

On account of the smallness of this value any variation in temperature of the thorium which might lead to a variation of gas pressure and gas quantity in the volume of the gauge will not affect the dissolved quantity q_s to any important extent. q_s can be considered as practically constant, and eqn. (14) can be checked by measuring the hydrogen pressure inside the gauge as a function of the temperature of the thorium, provided the gas is not sorbed appreciably by other parts of the gauge. It was found that such an additional sorption is negligible since at one particular temperature of the thorium the pressure in the gauge remained practically constant for periods which were much longer than those required for the measurement.

If $\log p$ is plotted against $1/T$ (fig. 13) it is seen that the experimental points lie sufficiently on a straight line to show that the exponential law of eqns. (14) is complied with. For the heat of solution H, which can be derived from the slope of such lines, an average value $H = -13{\cdot}5$ kcal was obtained (from three sets of measurements). This compares with a value $H_{\mathrm{mol}} = -22{\cdot}5$ kcal/mole which has been derived by Fowler and Smithells (1937) from measurements by Sieverts and Roell (1926). If our value is valid for an atomic gas it should be equal to $\frac{1}{2}H_{\mathrm{mol}}$. In view of the fact that Sieverts and Roell's value has to be based on three measured points only, the agreement appears to be as good as it can be expected.

§ 9. THE SORBED QUANTITIES AND THEIR DEPENDENCE ON TEMPERATURE

Sorbed quantities can be ascertained from eqn. (3) or (15) with sufficient accuracy provided the rate of sorption S does not become comparable with the sorption rate of the ionization gauges employed. The latter rate, on account of its inconstancy, cannot be measured very accurately. Consequently, if this rate is of the same order as the rate of sorption determined from eqn. (1), a certain error must be tolerated when evaluating the integral in (3) or (15).

The difficulty in question mainly arises if we attempt to ascertain the total quantity of gas which can be sorbed at one particular temperature. In order to determine this quantity (the capacity of sorption) the appropriate decay curve is to be measured until the sorption rate S of the thorium is equal to zero. As seen from figs. 3–6, $S=0$ is approached very gradually, and the error due to the uncertain value of the sorption rate of the ionization gauge becomes correspondingly great. Owing to this difficulty the sorbed quantities have only been calculated for a period during which a certain minimum value of rate of sorption was reached (100 cm³/sec for oxygen and 20 cm³/sec for hydrogen).

Table 4 shows that the sorbed quantity of oxygen increases with temperature considerably, more than thirtyfold from 350 to 950°K. It will be interesting to compare the values quoted with the quantity of gas which can be adsorbed on the surface of the thorium.

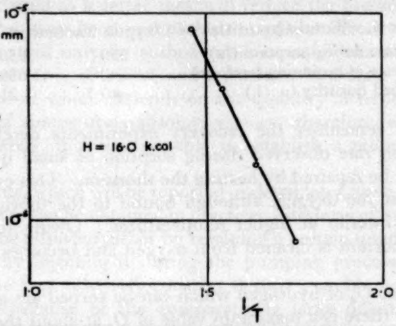


Fig. 13. **Hydrogen pressure above thorium as a function of temperature of thorium.**

If N_s denotes the number of sorbed gas atoms, we have for a diatomic gas at temperature T

$$N_s = 7{\cdot}08 \times 10^{16} \frac{273}{T} Q_s \qquad \ldots\ldots(18)$$

(Q_s in $\mu \times$ l.). On the other hand, if A denotes the geometric surface area of the thorium, fA the real adsorbing area and s the spacing of adsorbed atoms, the number of such adsorbed atoms is

$$N_{ad} = fA/s^2. \qquad \ldots\ldots(19)$$

For $N_{ad} = N_s$ and $T = 300$°K we have

$$f = 6{\cdot}45 \times 10^{16} s^2 Q_s / A. \qquad \ldots\ldots(20)$$

Assuming the formation of thorium dioxide during chemisorption, s is equal to half the lattice spacing of ThO_2 ($s = \frac{1}{2} \times 5{\cdot}59$ Å). Hence, when using $Q_s = 0{\cdot}7$ for $T = 350$°K in table 4 and $A = 4$ cm² we find $f = 8{\cdot}8$.

Such a ratio between real, adsorbing area and geometric area appears reasonable. The conclusion, therefore, seems to be justified that at low temperature oxygen is sorbed chemically at the surface of the thorium only.

If the much higher values of Q_s which are measured at $T = 950$°K, are used instead, a value $f = 330$ is obtained which appears rather excessive. One will therefore conclude that at higher temperatures not only the surface but also parts of the thorium lattice lower down participate in the process of sorption. Diffusion from the surface into the interior will take some part in the process.

On the other hand, it can be shown that this diffusion is rather limited, since no complete oxidation of the thorium can be obtained at the pressures and temperatures concerned. For such complete oxidation if M is the molecular weight of the gas and A_w the atomic weight of the thorium M/A_w mg gas/mg thorium should be sorbed. Since

$$1 \text{ mg gas} = \frac{1{\cdot}70 \times 10^4}{M} \frac{T}{273} \mu \times \text{l.} \qquad \ldots\ldots(21)$$

we find for $T = 300°\text{K}$ that 1 mg thorium can take up theoretically $1{\cdot}70 \times 10^4 \times 1{\cdot}1/A_w = 80{\cdot}5\mu \times \text{l.}$ oxygen ($A_w = 232{\cdot}1$). The 26 $\mu \times$ l. quoted in the last line of table 4 refer to 10 mg thorium and therefore to a degree of oxidation of only $2{\cdot}6/80{\cdot}5 = 3{\cdot}2\%$.

Table 4. Total Quantities of Oxygen Sorbed in Thorium

Temperature during sorption (°K)	350	950	950
Sorption rate at beginning (cm³/sec)	560	650	970
Total sorbed quantity ($\mu \times$ l.)	0·7	21	26

Let us now remember the recovery experiments described in §5. The decay in sorption rate observed during sorption of small quantities of oxygen ($\simeq 1\mu \times$ l.) could be repaired by heating the thorium. This can also be explained by assuming that the oxygen, although bound to the thorium chemically, can diffuse into the interior at higher temperatures. Owing to this diffusion the surface of the thorium is cleaned from oxygen and becomes capable of sorbing new quantities.

The quantities Q_s of hydrogen which can be sorbed are given in table 5. It will be seen that there is a maximum value of Q_s at about the same temperature as that observed for the maximum sorption rate S (at about 700°K). The fall in Q_s with rising temperature is to be expected on the basis of equations (10) or (12). In order to explain the maximum and the fall with decreasing temperatures beyond it we must realize that the quantity of gas dissolved in a metal does not only depend on the solubility of the metal but also on the extent to which the gas can diffuse into the metal. If there is no diffusion, no gas can be dissolved in spite of a high theoretical solubility. Since diffusion decreases with temperature by an exponential law, diffusion of hydrogen into the thorium will be more and more limited with falling temperature. The hydrogen will therefore be less and less able to reach the deeper parts of the thorium and the total quantity of hydrogen dissolved will consequently decrease in spite of an increase in the quantity q_s dissolved per unit weight according to equations (10) or (12).

Table 5. Total Quantities of Hydrogen Sorbed in Thorium

Temperature during sorption (°K)	500	565	630	740	845	935
Total sorbed quantity ($\mu \times$ l.)	1·08	1·42	2·12	1·98	0·73	0·34

§10. CONCLUSIONS

The experiments have shown that oxygen is sorbed chemically by thorium at all temperatures investigated, forming a stable oxide. The equilibrium pressure of oxygen above this oxide is immeasurably small. Provided the capacity of the thorium is not exhausted the rate of sorption for oxygen injected from outside is and remains high even at the lowest pressures accessible to measurement ($\simeq 10^{-7}$ mm Hg).

Hydrogen, however, is dissolved in the thorium and can be removed from it by heating. In equilibrium the hydrogen pressure outside the thorium is a function of the quantity of hydrogen dissolved inside. The rate of sorption is zero under these equilibrium conditions.

These results open up new possibilities for the use of getter materials employed for improving the gas atmosphere of electronic valves and tubes (Wagener 1952). The alkali earth metals, e.g. barium, which are normally used as such getters reduce the pressure of all residual gases existing in the valve or tube, independent of their nature. It is known, however, that the gases containing oxygen are mainly obnoxious while for instance hydrogen, provided its pressure is not too high, is not harmful and may even be beneficial in some cases. If materials such as thorium are used as a getter they will reduce the pressure of oxygen or carbon dioxide to the same level and with even greater efficiency than barium. Hydrogen, however, will only be sorbed until an equilibrium between inside and outside of the getter is obtained, and a constant hydrogen pressure will be established, the level of which depends on the quantity of hydrogen present in the tube. Instead of aiming at an absolute vacuum, therefore, which can hardly be obtained completely, it seems possible to establish a reducing atmosphere inside the tube.

Some experiments, which will be reported in detail later, have shown that such a reducing atmosphere can be obtained in practice. The level of hydrogen pressure to be aimed at can be adjusted either by retaining a certain quantity of hydrogen in the thorium or by injecting it during the pumping process. The pressure level attained can be kept relatively constant despite possible losses of hydrogen to glass walls etc. because of the comparatively large quantity of hydrogen dissolved in the getter. According to eqn. (17) which refers to conditions existing in practice, about 1000 times more hydrogen is dissolved in the thorium than exists outside. The thorium therefore acts as a reservoir for hydrogen from which losses can be easily replaced.

ACKNOWLEDGMENTS

Acknowledgment is made to the Engineer-in-Chief of the General Post Office for permission to make use of the information contained in this paper. The author also wishes to express his appreciation to Mr. R. E. Thorne for his assistance in conducting the experiments.

REFERENCES

ESPE, W., 1948, *Powder Metall. Bull.*, **3**, 100.

FOWLER, R. H., and SMITHELLS, J. C., 1937, *Proc. Roy. Soc.* A, **160,** 37.

GLASSTONE, S., 1940, *Physical Chemistry* (London : Macmillan), p. 850.

KIRSCHFELD, L., and SIEVERTS, A., 1929, *Z. phys. Chem.* A, **145,** 227.

LANGMUIR, I., 1950, *Phenomena, Atoms and Molecules* (New York : Philosophical Library), p. 150.

SIEVERTS, A., and ROELL, E., 1926, *Z. anorg. Chem.*, **153,** 298.

ULICH, H., 1948, *Physikalische Chemie*, Dresden, Leipzig, p. 137.

WAGENER, S., 1950, *Brit. J. Appl. Phys.*, **1,** 225; 1951, *Ibid.*, **2,** 132; 1952, *Proc. Instn. Elect. Engrs.*, Pt. III, **99,** 135.

Interferometric Studies of the Growth of Stearic Acid Crystals and their Optical Properties

By AJIT RAM VERMA and P. M. REYNOLDS
Royal Holloway College, University of London

Communicated by S. Tolansky; MS. received 26th November 1952, and in amended form 2nd February 1953

Abstract. On the basal planes of stearic acid crystals, grown from solution, growth spirals observed by optical and phase contrast microscopic techniques are reported. These growth spirals are in accordance with the theory of Burton, Cabrera and Frank. For the measurement of spiral step-heights multiple-beam interference methods were used, including a modified multiple-beam interference method, which is described. The minimum step-height found was 46·3 ± 0·8 Å, which agrees with the x-ray repeat unit. In addition, step-heights which were integral, as well as half integral, multiples of this unit were found. By interferometric methods the refractive index was found to be 1·42 ± 0·04 and for the two types of crystal rhomb observed the birefringence for directions of vibration parallel to the 'a' and 'b' axes was approximately 0·018 and 0·015 respectively for $\lambda = 5461$ Å.

§ 1. INTRODUCTION

SINCE Burton, Cabrera and Frank (1949, 1951) proposed a screw dislocation mechanism for the growth of crystals from vapour, observations of the related growth spirals have been reported on several ionic, homopolar, molecular and metallic crystals. Among these observations measurements of growth spirals with step-heights equal to a parameter of the x-ray unit cell have been reported on beryl (Griffin 1950), the long chain paraffins (Dawson and Vand 1951, Dawson 1952), silicon carbide (Verma 1951 a, b, Amelinckx 1951), and haematite crystals (Verma 1952). On cadmium iodide (Forty 1952) all the measured spiral step-heights were found to be multiples of the x-ray unit, whilst on other crystals, e.g. gold (Amelinckx 1952), the existence of dislocations of multiple strength has been inferred from the visibility of the steps. The only crystal upon which dislocations of both unit and multiple strength have been measured is silicon carbide.

In previous observations upon the growth of long chain organic molecules only an electron microscopic technique has been applied. The spiral step-heights were found to be only equal to the x-ray unit. In the present study upon the growth of crystals of the long chain molecule stearic acid [$CH_3(CH_2)_{16}COOH$], optical and interferometric techniques have been applied. A preliminary account has already been given elsewhere (Reynolds and Verma 1953). The following is a detailed account, in part I of the growth properties, and in part II of the optical properties, of small, nearly perfect crystals.

PART I

§ 2. PREPARATION OF STEARIC ACID CRYSTALS

The crystals were grown from a dilute benzene solution of stearic acid (B.D.H. pure laboratory reagent, concentration 0·3% by weight). Thin crystal plates formed if a small drop of this solution was placed upon a cooled glass flat and the rate of evaporation suitably restricted by enclosing it. At higher rates of evaporation dendritic chains of small crystals were obtained, whilst if the droplet evaporated rapidly from a warmed glass plate, features resembling circular spirals were obtained (fig. 1)*. An examination of a line section across these features by the dark slit profile microscope (Tolansky 1952) has shown that the edges of the 'circular spirals' are ridges of material, and that the surface between successive turns of the spiral is not plane, but has an appreciable curvature. These are, therefore, 'evaporation figures', and are to be distinguished from growth spirals.

When the crystals were grown upon a silvered glass surface for interferometric examination it was found that the solvent, benzene, did not appreciably increase the absorption of the silver layer.

§ 3. CRYSTAL STRUCTURE

Stearic acid is polymorphic (Piper, Malkin and Austin 1926) and forms monoclinic prismatic crystals. The crystal molecule has a chain of carbon atoms arranged in a zigzag manner, the number of carbon atoms in this chain being equal to that in the chemical molecule. The chain axes all lie in the symmetry plane (*ac* plane), parallel to the *c*-axis and inclined to the basal plane (*ab* plane) at an angle β. Müller (1927) reported one form with lattice parameters $a=5{\cdot}546$ Å, $b=7{\cdot}381$ Å, $c=48{\cdot}84$ Å, $\beta=63^\circ 38'$ so that $c \sin\beta=43{\cdot}76$ Å. Other forms have been reported with the longer layer spacings equal to 46·6 Å, 43·75 Å, 39·75 Å (Francis, Collins and Piper 1937), 41·5 Å (Thibaud and Dupré la Tour 1930).

§ 4. MICROSCOPIC OBSERVATIONS OF GROWTH FEATURES

For the observation of growth features the crystals were silvered by thermal evaporation under high vacuum, and examined with bright field illumination, or with a positive phase contrast equipment. All the growth spirals were observed on the basal planes of the crystals, but there were many crystals upon which there was no evidence of any surface structure. Numerous examples of growth features originating from one and a larger number of screw dislocations were observed, including the interaction of their growth fronts (e.g. formation of closed loops in fig. 7). Their behaviour is similar to features already reported on other crystals, therefore observations were concentrated upon growth features which are peculiar to stearic acid.

The observed growth spirals could be divided into two types, firstly, rectilinear spirals having the symmetry of the crystal face and secondly curvilinear spirals.

Figure 2 shows a spiral of the first type. In such a rectilinear spiral the growth edge is parallel to the edges of the crystal plate which correspond to the close-packed [110] directions. This is shown diagrammatically in fig. 3, which

* For figs. 1, 2, 4, 5, 6, 7, 8 see Plate.

represents the projection of a layer of long chain molecules on the (*ab*) plane (neglecting the inclination β). Figure 4 is another example. In fig. 2 the successive turns of the growth edge are more widely spaced towards the centre yet remain parallel to the edges of the crystal plate. This indicates that the last phase of growth occurred under lower supersaturation or at a higher temperature. The final section of the growth edge is almost parallel to the *a*-axis, which after the [110] is the next most close-packed direction. It is probable, therefore, that some movement of the dislocation occurred after growth had ceased.

In the second type of spiral the obtuse, and also sometimes the acute, angle between the growth edges became rounded, giving the spiral a 'leaf' shape. Opposite diagonals of these spirals were not generally collinear but were inclined at an angle usually less than 10° (see fig. 5). Numerous growth spirals were observed in which the spiral step-edges were not only curved but also inclined at a fairly large angle to the boundary of the crystal plate, which indicates that although growth spirals may take the symmetry of the crystal face, they do not necessarily control it.

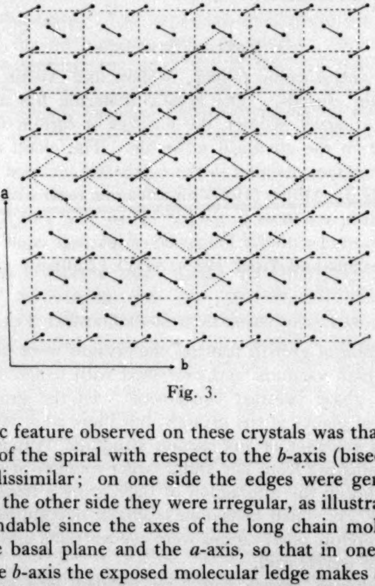


Fig. 3.

A characteristic feature observed on these crystals was that the growth edges on opposite sides of the spiral with respect to the *b*-axis (bisectrix of acute angle of rhomb) were dissimilar; on one side the edges were generally regular and smooth, whilst on the other side they were irregular, as illustrated in fig. 5. This effect is understandable since the axes of the long chain molecules are inclined at an angle to the basal plane and the *a*-axis, so that in one half of the crystal with respect to the *b*-axis the exposed molecular ledge makes an acute angle with the basal plane, whilst in the other half of the crystal this angle becomes obtuse. The van der Waals binding forces between neighbouring chains are therefore not the same for every growth edge in the [110] directions, and preferential etch may occur.

The visibility of the multimolecular steps on stearic acid was markedly less than that of steps of similar heights on silicon-carbide crystals, in which the edges are very much more regular.

In almost every case it was observed that the point of emergence of the dislocation was at the centre of the crystal plate; this is to be expected in an isolated single crystal since a dislocation line is effectively under tension (Frank 1949). The point of emergence of the dislocation was occasionally marked by a regular dot which is probably due to the nucleation of the crystal by an impurity.

The existence of oriented overgrowth on certain crystal plates was also established.

§ 5. INTERFEROMETRIC METHODS

(i) *Multiple-Beam Fizeau Fringes*

An external optical flat could only be used with a few of the larger crystals on which the growth steps could be clearly resolved at magnifications less than × 200. Using multiple-beam Fizeau fringes (Tolansky 1948), it was convenient to adjust the tilt of the silvered optical flat until the fringes contoured the edges of the growth spiral. Assuming that the steps were all of equal height h, the step-height is then given by

$$h=\lambda/2n_1, \qquad \ldots\ldots(1)$$

where λ is the wavelength of the monochromatic light, and n_1 (not necessarily integral) the number of growth steps between the intensity peaks of two successive fringes. Entry 1 in table 1 is an example of the application of this method, and is the most accurate value obtained in this study.

Since the crystals were usually small it was necessary to examine them at higher magnifications. A thin film technique (Tolansky and Omar 1952) was therefore used which indicated that certain step-heights were multiples of the x-ray repeat unit.

(ii) *Multiple-Beam Internal Interference Fringes*

For a more accurate determination of step-height interference fringes formed by internal reflection between the two surfaces of the crystal were used. Forty (1952) has used the two-beam internal interference fringes for CdI_2 crystals. This method has been improved utilizing multiple-beam fringes by silvering opposite faces of the crystal, a method similar to that used by Tolansky (1948) for mica. Therefore the crystals of stearic acid were grown on a silvered optical flat ($\lambda/40$), and their upper surfaces were also silvered. It was found that the crystal surface in contact with the flat did not generally acquire any growth features, and spirals only developed on the exposed surface; only such crystals were selected for the measurement of step-height. The Fizeau fringes formed between the flat base and 'vicinal faces' of any growth pyramids contour the growth steps and, if these fringes are sufficiently sharp, appear double, due to the birefringence of the crystal (see fig. 6), a feature which cannot be seen in two-beam fringes. Formula (1) now becomes

$$h=\lambda/2\mu n_2, \qquad \ldots\ldots(2)$$

where n_2 is the number of steps between any two successive fringes having the same plane of vibration, and μ is the corresponding refractive index for light of wavelength λ. Since the peak of the Fizeau fringe may not occur on a step along a selected radial line n_2 will not necessarily be integral. It is then necessary to add to (or subtract from) the integral number of steps between two successive fringes a small fraction, as done by Forty (1952), However, multiple-beam fringes have shown that the surface of the stearic acid crystal is not perfect to within a few ångströms.

§6. Discussion of the Measured Step-Heights

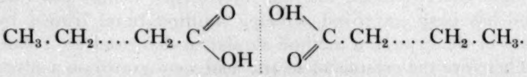
A large number of measurements of step-heights have been made; of these the most accurate and important ones are given in table 1. Many higher steps have been measured, but in these cases, for step-heights greater than four times the X-ray repeat unit, the error in neglecting the fractional part of n_2 is greater than 10%, and it cannot be established whether these steps are an integral multiple of a unit.

Table 1

	n	μ	h (Å)	ψ	Multiple	Remarks
1	$58 < n < 60$	1·0	$46{\cdot}3 \pm 0{\cdot}8$	—	$1 \times 46{\cdot}3$	Air film
2	$29 < n < 30$	1·42	$65{\cdot}2 \pm 0{\cdot}9$	74°	$1\frac{1}{2} \times (43{\cdot}5 \pm 0{\cdot}6)$	
					$2 \times (49{\cdot}3 \pm 1{\cdot}2)$	
					$2\frac{1}{2} \times (39{\cdot}5 \pm 1)$	
					$3 \times (32{\cdot}9 \pm 0{\cdot}8)$	
3	$19 < n < 20$	1·42	$98{\cdot}7 \pm 2{\cdot}5$	74°	$2 \times (46{\cdot}5 \pm 4)$	
4	$32 < 2n < 33$	1·42	118 ± 2	74°	$2\frac{1}{2} \times (47{\cdot}2 \pm 1)$	Fig. 4
					$3 \times (39{\cdot}3 \pm 1)$	
5	$20 < n < 21$	1·0	133 ± 3	—	$3 \times (44{\cdot}3 \pm 1)$	Air film
6	$10 < n < 11$	1·42	183 ± 9	74°	$4 \times (45{\cdot}7 \pm 2{\cdot}3)$	
7			196	—	4×49; $4\frac{1}{2} \times 43{\cdot}6$	Fringes of equal chromatic order
					$5 \times 39{\cdot}2$	

The step-heights in table 1 may be divided into three types: (i) those equal to the X-ray unit, (ii) a multiple, and (iii) an odd half-integral multiple of this unit.

The occurrence of step-heights equal to an integral multiple (including unity) of the X-ray repeat unit are readily understood as follows: in stearic acid there is an active carboxyl group at the end of the molecular chain which results in the association of the molecules in pairs and gives rise to a double-layer crystal structure as shown:



$$\mathrm{CH_3.CH_2\ldots.CH_2.C}\overset{\displaystyle /\!\!/ O}{\underset{\displaystyle \backslash OH}{}}\quad\vdots\quad\overset{\displaystyle OH\backslash}{\underset{\displaystyle O/\!\!/}{}}\mathrm{C.CH_2\ldots.CH_2.CH_3.}$$

Although the binding forces between the carbon atoms in a chain are homopolar, the parallel chains are bound laterally by relatively weak van der Waals forces. Dislocations of integral multiple strength are therefore readily created by the slipping of long chain molecules past one another.

The most interesting step-heights are entries 2, 3, 4, and 7 in table 1, which may be interpreted either as integral multiples of different polymorphs or as half-integral multiples of a known X-ray unit. But in entry 2 the step-height is

$$(65{\cdot}2 \pm 0{\cdot}9)\,\text{Å} = 1\tfrac{1}{2}(43{\cdot}5 \pm 0{\cdot}6)\,\text{Å} = 2(32{\cdot}6 \pm 0{\cdot}45)\,\text{Å}.$$

No polymorph of stearic acid with long layer spacing equal to 32·6 Å is known, and it is unlikely for a fatty acid with carbon content C_{18}. Furthermore it has been observed that stearic acid crystals grown from benzene always had the long layer spacing of 43·75 Å (Francis, Collins and Piper 1937), which therefore supports the conclusion that the step-heights are odd half-integral multiples of the X-ray unit. It may be interpreted, then, that under certain circumstances single molecules can occur without pairing, but so far no step-height equal to half the X-ray unit has been found.

PART II

§ 7. OPTICAL PROPERTIES

Stearic acid, which is monoclinic, will be expected to be an optically biaxial crystal. Of the three directions of vibration of the optical ellipsoid, one coincides with the symmetry axis *b*, the other two lying in the symmetry plane *ac* (Thibaud and Dupré la Tour 1930). When examined between crossed nicols, the rhomb-shaped crystals show mutually perpendicular extinction directions.

The refractive index of a small crystal (size $\sim 200\,\mu$) has been determined interferometrically: comparing eqns. (1) and (2), $\mu = n_1/n_2$. In one growth feature (entry 1 in table 1) it was found $59 < n_1 < 60$ and the corresponding internal interference fringes gave $41 < n_2 < 42$, from which $\mu = 1{\cdot}42 \pm 0{\cdot}04$.

Two types of crystal rhombs of stearic acid were observed: (i) with the acute angle ψ between [110] and [1$\bar{1}$0] measured in the (001) plane equal to 74°, and (ii) with this acute angle ψ equal to 56° (fig. 7) These crystals differ in their optical properties.

(i) $\psi = 74°$.

The refractive index for a vibration parallel to the *b*-axis (called μ_b) is less than μ_a. This is concluded from the doubled Fizeau fringes (fig. 6) in which the *b* vibration corresponds to the component nearer to the centre of the growth hill, and hence to a larger interferometric gap and a smaller refractive index. It is confirmed by fringes of equal chromatic order in which the vibration parallel to *b* (marked as *b* in fig. 8) is the component which is towards the shorter wavelength end of the spectrum. This is to be expected since the *b* vibration is perpendicular to the chain length, and should therefore give the minimum refractive index.

(ii) $\psi = 56°$.

In contrast to the above, the refractive index for the vibrations parallel to the acute bisectrix is greater than for the vibration perpendicular to it.

The birefringence $d\mu$ is obtained directly from the fringes of equal chromatic order and, as shown by Tolansky (1948), is given by $d\mu/\mu = d\lambda/\lambda$, where $d\lambda$ is the wavelength separation between the two components of the same order. In table 2 the fractional birefringence $d\lambda/\lambda$ (for the shorter wavelength vibration) and an approximate value of the birefringence $d\mu$ is given for the two types of crystal. It was assumed that $\mu = 1{\cdot}42$.

Table 2

Crystal type	λ (Å)	$d\lambda/\lambda$	$d\mu$
$\psi = 74°$	5703	0·013	0·0185
	5614	0·014	0·020
	5548	0·013	0·0185
	4700	0·015	0·021
	4543	0·013	0·0185
$\psi = 56°$	5870	0·010	0·014
	4425	0·011	0·016

It can be calculated from the fringes of equal chromatic order shown in fig. 8 that the thickness of the crystal is of the order of five wavelengths of green light, and in a few cases we have measured the birefringence of even thinner crystals, but then the fringe width in terms of wavelengths is increased, so that the accuracy decreases. It is interesting to compare these values of birefringence with the value of $1{\cdot}535-1{\cdot}510=0{\cdot}025$ reported by Thibaud and Dupré la Tour (1930).

§ 8. CONCLUSION

It is shown that stearic acid crystals can grow in the form of a spiral originating from a screw dislocation. Multiple-beam interference techniques have been applied which show that the spiral step-heights are equal to integral or half-integral multiples of the X-ray repeat unit, within the limits of experimental error. These interferometric techniques have also given the refractive index and birefringence of crystals having a thickness of only a few wavelengths of light.

ACKNOWLEDGMENTS

We have great pleasure in thanking Professor S. Tolansky for his kind interest and for the facilities of his laboratory during this work. One of us (A. R. V.) wishes to take this opportunity of expressing his thanks for the Imperial Chemical Industries Fellowship of the University of London, and to the University of Delhi for an extension of study leave.

REFERENCES

AMELINCKX, S., 1951, *Nature, Lond.*, **167,** 939; 1952, *C. R. Acad. Sci., Paris*, **234,** 113.
BURTON, W. K., CABRERA, N., and FRANK, F. C., 1949, *Nature, Lond.*, **163,** 398; 1951, *Phil. Trans. Roy. Soc.* A, **243,** 299.
DAWSON, I. M., 1952, *Proc. Roy. Soc.* A, **214,** 72.
DAWSON, I. M., and VAND, V., 1951, *Proc. Roy. Soc.* A, **206,** 555.
FORTY, A. J., 1952, *Phil. Mag.*, **43,** 72, 377.
FRANCIS, F., COLLINS, F. J. E., and PIPER, S. H., 1937, *Proc. Roy. Soc.* A, **158,** 691.
FRANK, F. C., 1949, *Discussions of the Faraday Society*, No. 5, 76.
GRIFFIN, L. J., 1950, *Phil. Mag.*, **41,** 196.
MÜLLER, A., 1927, *Proc. Roy. Soc.* A, **114,** 542.
PIPER, S. H., MALKIN, T., and AUSTIN, H. E., 1926, *J. Chem. Soc.*, 2310.
REYNOLDS, P. M., and VERMA, A. R., 1953, *Nature, Lond.*, **171,** 486.
THIBAUD, J., and DUPRÉ LA TOUR, F., 1930, *C. R. Acad. Sci., Paris*, **191,** 200.
TOLANSKY, S., 1948, *Multiple-Beam Interferometry of Surfaces and Films* (Oxford: University Press); 1952, *Nature, Lond.*, **169,** 445.
TOLANSKY, S., and OMAR, M., 1952, *Nature, Lond.*, **170,** 81.
VERMA, A. R., 1951 a, *Phil. Mag.*, **42,** 1005; 1951 b, *Nature, Lond.*, **167,** 939; 1952, *Proc. Phys. Soc.* B, **65,** 806.

RESEARCH NOTES

Direct Measurement of the Specific Heat at Constant Volume of Pentane and Alcohol

By M. O. Bryant and G. O. Jones
Department of Physics, Queen Mary College, University of London

MS. received 30th January 1953

No direct measurements of c_v appear to have been made for liquids other than the liquefied 'permanent gases', because of the difficulty of containing them at constant volume. Instead, measurements are usually made of c_p, and c_v calculated through the equation $c_p - c_v = TV\alpha^2/\kappa$, where α and κ are the expansivity and compressibility respectively. Since c_p is normally measured only at 1 atm pressure this gives only values of c_v at 1 atm, and it is necessary to use, say, $(\partial c_v/\partial V)_T = T(\partial^2 p/\partial T^2)_v$ in order to obtain a true c_v. The application of these formulae requires a full knowledge of the equation of state data. In liquids, where c_p/c_v may be 1·5 or greater, the use of existing state data in this way is quite inadequate to determine whether there is any significant difference in form between the curves of c_p and c_v against temperature—apart from their relative displacement.

We have re-examined the question of making accurate direct measurements of the quantity c_v because of its importance in statistical-mechanical theories. With a suitably designed calorimeter the actual thermal expansion of the liquid contained may be reduced at least to 5–10% of that corresponding to its expansivity at 1 atm, so that c_v as measured will be systematically high by only 2–3%. A correction for this may safely be applied even if κ is known only approximately, a 10% error in κ leading to one of only 0·5% in c_v. Also, because α/κ does not vary greatly from one liquid to another at corresponding temperatures, an accuracy of about 1% in c_v would be possible if only α had been independently measured.

Two liquids for which state data exist—*n*-pentane and ethyl alcohol (Bridgman 1913, Seitz and Lechner 1916)—have been investigated, using a thick-walled steel calorimeter surrounded by a radiation shield and contained in a vacuum chamber. Escape of liquid during measurements could be prevented by cooling the tube communicating with the calorimeter to a temperature below the freezing point of the liquid, after completely filling the calorimeter and tube. The quantity of liquid contained could be varied by releasing this 'freezing-valve' at differing temperatures after filling at a given temperature. The measurements of heat capacity were estimated to be accurate to about 0·25%, but because of the large heat capacity of the calorimeter itself—5 to 6 times that of the liquid contained—the final accuracy in c_v is estimated as about 1% at the lowest and 2% at the highest temperatures. The experimental uncertainty was here greater than that introduced by the necessity for the correction mentioned above, so that a somewhat lighter calorimeter could in fact have been used.

For both liquids c_p and c_v at 1 atm were measured over a range of about 100 centigrade degrees immediately above the freezing point, and for alcohol values of a true c_v were determined up to about 750 atm at the specific volume corresponding to a temperature just above the freezing point (figs. 1 and 2).

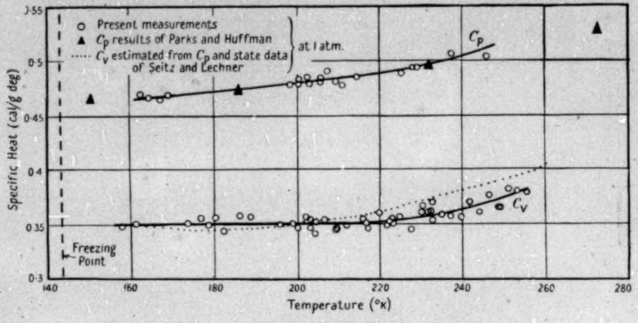


Fig. 1. Specific heats of liquid *n*-pentane.

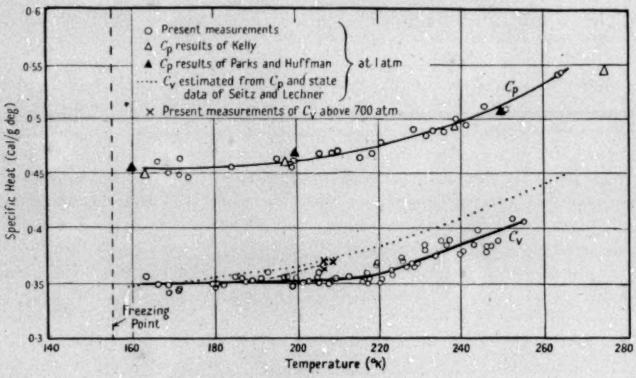


Fig. 2. Specific heats of liquid ethyl alcohol.

The results for c_p are in good agreement with those obtained in the careful measurements of Kelly (1929) and Parks and Huffman (1929, 1930) in spite of the much larger relative thermal capacity of the calorimeter in our experiments. The results for c_v are in fair agreement with those calculated from the values of c_p and the state data, and it will be seen that c_v increases with pressure, a result which could not have been certainly predicted from the somewhat conflicting state data at the higher pressures.

The direct method applied in this way appears to us to be quite suitable for the measurement of c_v for liquids. We have considered the possibility of introducing a solid inclusion having high expansivity into the calorimeter in order to reduce the expansion of the liquid to zero. The addition to the heat capacity which this would entail would, however, cause greater uncertainty in the final estimate of c_v, even in the most favourable case, than that introduced by the correction for expansion made as indicated above. We have considered also the possibility of making direct measurements of c_v for solids by the introduction of a suitable second inclusion and of a very small quantity of liquid for the transmission of pressure. In the most favourable cases the actual expansion cannot be reduced below about half that corresponding to the expansivity at 1 atm, so we conclude that such measurements are impracticable.

We are glad to acknowledge the encouragement and help of Professor H. R. Robinson, and the loan of apparatus by Imperial Chemical Industries Ltd.

References

Bridgman, P. W., 1913, *Proc. Amer. Acad.*, **49**, 1.
Kelly, K. K., 1929, *J. Amer. Chem. Soc.*, **51**, 779.
Parks, G. S., and Huffman, H. M., 1929, *J. Phys. Chem.*, **31**, 1842; 1930, *J. Amer. Chem. Soc.*, **52**, 4381.
Seitz, W., and Lechner, G., 1916, *Ann. Phys., Lpz.*, **49**, 93.

Application of the Thermodynamics of Irreversible Processes to the Theory of the Magnetron

By P. A. LINDSAY and G. D. SIMS
Research Laboratories, The General Electric Co., Ltd., Wembley

MS. received 11*th December* 1952

At least twice in the past (Richardson 1921, Gabor 1945), the thermodynamics of reversible processes (or rather thermostatics if we accept de Groot's definition) has been applied to the problem of thermionic valve operation. However, by its very definition, the thermodynamics of reversible processes can only be applied to systems in thermal equilibrium, i.e. characterized by a stationary value of entropy. In the case of magnetrons, this corresponds to zero current from the cathode to the anode of the valve leaving out completely the whole field of current conduction and oscillations.

It seems that the relatively new ' thermodynamics of irreversible processes ' as proposed by Onsager (1931), Casimir (1945) and Callen (1948) and finally described by de Groot (1951) would provide a sound theoretical foundation, of a very broad nature, on which to build a satisfactory theory of magnetron conduction and oscillations.

Briefly, two fundamental features of the thermodynamics of irreversible processes make it singularly suitable for an application to the magnetron problem : (i) Onsager's proof that the laws governing the irreversible processes of thermodynamics can be logically derived from the reversibility of microscopical phenomena, which in itself provides a very good basis for extending the techniques of statistical mechanics to the problem of conduction and oscillation in thermionic valves, and (ii) the fact that the macroscopic equations describing the systems under consideration go beyond the normal field of thermal equilibrium and cover the field of physical and chemical phenomena in which the entropy of the system does not remain constant (see Callen 1948, Sommerfeld and Frank 1931, de Groot 1951, p. 195ff. for some limiting cases). This again is very important in the field of thermionic valves, where in the past no such phenomena would have been considered from the thermodynamic point of view.

The broad character and power of the thermodynamic approach can be seen best from two examples. In the first case still using the old thermodynamics of reversible processes it is possible to show (Meixner 1942, de Groot 1951) that in an ideally cut-off magnetron (a macroscopically reversible system) the electron cloud must rotate as a solid mass round the cathode with constant angular velocity (Brillouin single stream solution), and secondly (de Groot 1951, p. 47ff.) using the new thermodynamics of irreversible processes it is possible to prove quite generally that the principle of reciprocity holds for arbitrary passive electrical networks.

The specification of the magnetron as a system which could be treated by the methods of the thermodynamics of irreversible processes requires some consideration of what is meant by temperature inside the valve. It seems best to define temperature in this case after Chapman and Cowling (1952, p. 37) by the relation

$$T=(\boldsymbol{m}/3\boldsymbol{k})\overline{(v-\bar{v})^2}$$

where $\bar{v}$ is the mean electron velocity, v the actual velocity, $\boldsymbol{k}$ Boltzmann's constant and $\boldsymbol{m}$ the mass of an electron. The averages are taken in this case over all existing electron velocities likely to occur at a given point in space. This temperature will be a continuously changing function throughout the electron cloud, whatever the electrical boundary conditions (cf. Hines 1951) and the various interaction effects between temperature and the normal electron flow within the valve will have an important bearing on the noise characteristics (see, for example, Cutler and Quate 1950).

With this in mind then (de Groot 1951, Callen 1948) we could set up equations of the form

$$\begin{bmatrix} J_1 \\ J_2 \\ J_3 \\ J_4 \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} & L_{13} & L_{14} \\ L_{12} & L_{22} & L_{14} & L_{24} \\ -L_{13} & -L_{14} & L_{11} & L_{12} \\ -L_{14} & -L_{24} & L_{12} & L_{22} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} \quad \ldots\ldots (1)$$

where J_1 and J_3 are components of the particle current density $-\mathbf{J}$, and J_2 and J_4 are components of the heat current density $\mathbf{Q}$. X_1 and X_3 are components of the electric field $(1/T)$ grad $(\boldsymbol{e}\phi)$ and X_2 and X_4 are components of the temperature gradient grad $(1/T)$ while the L_{ij} are constant coefficients. The ' forces ' and ' fluxes ' have been suitably chosen to obey the general equation for the generation of entropy (de Groot 1951)

$$d(\Delta S)/dt=\text{grad }(1/T)\,.\,\mathbf{Q}-(1/T)\text{ grad }\boldsymbol{e}\phi\,.\,\mathbf{J} \quad \ldots\ldots (2)$$

while in addition full use has been made of the Onsager reciprocal relations

$$L_{ij}(B)=L_{ji}(-B) \quad \ldots\ldots (3)$$

where B is the magnetic field.

Equations 1 and 3 are very important as they should allow us to draw quite general conclusions about the coefficients L_{ij} without having to solve the complicated equations for ϕ and T.

Finally the properties of the observed noise in thermionic valves or for that matter in any other thermodynamic system, are strictly determined by the laws of irreversible thermodynamics. The usual Nyquist relation occurs as a special case of a theorem due to Callen and Welton (1951) and later expanded by Callen and Greene (1952), which, particularized for the case in question, states that the mean square noise voltage $\overline{V^2}$ is given by

$$\overline{V^2} = \left(\frac{2}{\pi}\right) \boldsymbol{k}T \int \mathscr{R}\left(\frac{1}{Y(\omega)}\right) d\omega,$$

where $Y(\omega)$ is the admittance of the system, the spectral density being given by

$$-\left(\frac{2}{\pi}\right)^{1/2} \boldsymbol{k}\, \mathscr{R}\left(\frac{1}{(Y\omega)}\right) d\omega.$$

It is of interest to note that no reactive terms appear in any of these expressions.

REFERENCES

CALLEN, H. B., 1948, *Phys. Rev.*, **73,** 1349.
CALLEN, H. B., and GREENE, R. F., 1952, *Phys. Rev.*, **86,** 702.
CALLEN, H. B., and WELTON, T. A., 1951, *Phys. Rev.*, **83,** 34.
CASIMIR, H. B. G., 1945, *Rev. Mod. Phys.*, **17,** 343.
CHAPMAN, S., and COWLING, T. G., 1952, *The Mathematical Theory of Non-Uniform Gases* (Cambridge : University Press), p. 37.
CUTLER, C. C., and QUATE, C. F., 1950, *Phys. Rev.*, **80,** 875.
GABOR, D., 1945, *Proc. Roy. Soc.* A, **183,** 436.
DE GROOT, S. R., 1951, *Thermodynamics of Irreversible Processes* (Amsterdam : North Holland Publishing Co.).
HINES, M. E., 1951, *J. Appl. Phys.*, **22,** 1385.
MEIXNER, I., 1942, *Ann. Phys., Lpz.*, **41,** 409.
ONSAGER, L., 1931, *Phys. Rev.*, **38,** 2265.
RICHARDSON, O. W., 1921, *Emission of Electricity from Hot Bodies* (London : Longmans Green).
SOMMERFELD, A., and FRANK, N. H., 1931, *Rev. Mod. Phys.*, **3,** 1.

The Dependence of the Dielectric Strength of Pure Liquids on Cathode Material

BY T. J. LEWIS
Queen Mary College, University of London

MS. received 29th January 1953

INVESTIGATIONS of the dielectric strength of simple highly purified organic liquids (Dornte 1939, Edwards 1951, Salvage 1951 for example) have shown that the strength is of the order of 10^6 v/cm for most of these liquids. In such average fields it has been considered likely that field emission of electrons from the cathode occurs, since the fields there may be several times greater than the average value due to surface irregularities, space charge formation, and the effective lowering of the energy barrier by the presence of the dielectric liquid (Le Page and Du Bridge 1940). If, as seems plausible, such emission in adequate amounts initiates the breakdown process, then it might be expected that the dielectric strength would depend on the nature of the cathode surface and in particular on its work function.

Results in general however (e.g. Sorge 1924, Kronig and Van de Vooren 1942, Salvage 1951) show a rather irregular dependence on the cathode material. The present author (Lewis 1953) using spherical electrodes, producing uniform fields, in pure *n*-hexane found the strength to be practically independent of the metal of the cathode. It must be remembered of course that surface contaminations of various sorts may mask the true effect. There is, however, apart from unknown surface conditions, a further factor to be considered in the breakdown of the liquid which may serve to obscure any cathode dependence. Even if it is assumed that the fields just prior to breakdown are enough to produce a sufficient supply of electrons from the cathode, there will need to be a further process of electron multiplication in the bulk of the liquid in such a way that breakdown ensues in a catastrophic manner.

The principal requirement of this latter process is likely to be ionization of the liquid molecules themselves by electrons of sufficient energy. This requires the setting up of a field in the liquid such that electrons, in spite of retarding collisions with the liquid molecules, may accelerate and acquire ionizing energies. A possible mechanism by which the electrons lose energy in a liquid has been outlined by von Hippel (1937).

Thus with this picture the dielectric strength will be determined by both factors: a cathode field sufficient for electron emission and a field in the bulk of the liquid to produce adequate ionization.

Experimental evidence suggests that these two fields are of the same order of magnitude in most pure insulating liquids. If, for a particular liquid and electrode configuration, the field in the liquid required tc promote ionization is such that the corresponding cathode field is producing a copious emission of electrons then, assuming the argument above, it is likely that little dependence on cathode material will be found. The breakdown is then influenced mainly by the characteristics of the liquid. On the other hand, cathode dependence might be expected if the ionization field in the liquid is lower, and the process of breakdown determined by the cathode emission reaching suitable proportions.

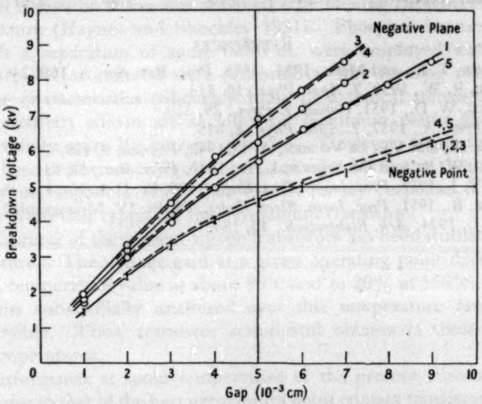
The normal measurement of the dielectric strength of very pure liquids using spherical electrodes giving uniform fields may show no consistent dependence on the cathode work function because the arrangement tends to give conditions in the first category. The position is different if a non-uniform field, produced by a point–plane electrode arrangement is used. For such a system the field at the point is well in excess of that at the plane, provided the point radius is small enough, and the non-uniformity increases as the gap increases. Such a non-uniform field is likely to produce a polarity effect resulting in different breakdown voltages for a negative plane and a negative point. Two main reasons might be suggested for this.

Firstly the mechanism suggested above would be likely to produce the effect. If the plane is made the cathode, then the field there will tend to be lower than in the liquid and the latter will be overstressed before the cathode emission becomes sufficient, i.e. a cathode dependence might be expected. With the point negative, copious emission is very probable before the field in the bulk of the liquid is great enough for breakdown, and in this case the breakdown is largely governed by the liquid itself.

A second possibility will be concerned essentially with ionization processes in the bulk of the liquid itself. Since positive ions have low mobilities and are inefficient ionizing agents compared with electrons, there would be a difference between the breakdown processes occurring when the point is positive and when it is negative. In air, it has been shown (Allibone and Meek 1938) that breakdown occurs at higher voltage when the point rather than the plane is negative and it would be plausible to consider that the conditions would be similar in a liquid.

The author has investigated the behaviour suggested above, using a point–plane electrode configuration of various metals in liquid *n*-hexane. The electrodes, free of grease and surface scratches, were placed in a cell into which the carefully dried and filtered liquid was then distilled. Dielectric strength measurements were made using direct voltage. The technique of measurement and the various cleaning processes have been described elsewhere (Lewis 1953).

The figure shows the breakdown strength as a function of gap width for the various electrodes tested; the coefficient of variation of results for any one test was about 1% and individual curves could be reproduced to within 2%. In each case the radius of the point was approximately $6{\cdot}5 \times 10^{-3}$ cm.



Breakdown voltage of point–plane electrodes in *n*-hexane.

(1) Al point and plane. (3) Cr point and plane.

(2) Cu point and plane. (4) Al point and Cr plane.

(5) Cr point and Al plane.

The dependence on cathode material is slight when the point is negative, but when the plane is negative a marked dependence is found, especially at the larger gaps, and suggests that the material of higher work function gives a higher breakdown voltage. For the smaller and therefore more uniform gaps, the difference begins to disappear. The negative plane curves lie above those for the negative point which is to be expected for a process initiated at the cathode. The evidence given by the measurements using a cathode and an anode of different materials (see figure) further indicates this and shows that the anode material plays little part in the breakdown process.

Approximate calculations for a gap of 5×10^{-3} cm show that for aluminium and chromium electrodes the field strengths at breakdown are as shown in the table. These figures, in spite of the difficulty of measuring the radius of the point and of knowing the field distortions occurring near the point, do indicate that with the plane negative the liquid is stressed to much higher values than those occurring with negative point breakdown long before the breakdown field is reached. Breakdown does not occur under these higher stresses presumably because of the dearth of electrons from the plane cathode.

It is suggested that the results obtained previously with uniform fields (Lewis 1953) correspond roughly to the present negative point breakdown where cathode dependence is slight. It seems possible therefore that a strong correlation between dielectric breakdown voltage and cathode material will be found only in suitable non-uniform fields or perhaps in liquids having a comparatively low dielectric strength.

Table

	Field at plane (mv/cm)		Field at point (mv/cm)	
	Al	Cr	Al	Cr
Plane Negative	0·92	1·11	1·85	2·24
Point Negative	0·74	0·75	1·49	1·51

References

Allibone, T. E., and Meek, J. M., 1938, *Proc. Roy. Soc.* A, **169,** 246.
Dornte, R. W., 1939, *J. Appl. Phys.*, **10,** 514.
Edwards, W. D., 1951, *Canad. J. Phys.*, **29,** 310.
von Hippel, A., 1937, *J. Appl. Phys.*, **8,** 815.
Kronig, R., and Van de Vooren, A. I., 1942, *Physica*, **9,** 139.
Le Page, W. R., and Du Bridge, L. A., 1940, *Phys. Rev.*, **58,** 61.
Lewis, T. J., 1953, *Proc. Instn. Elect. Engrs.*, **100,** Pt. II A, No. 3, in the press.
Salvage, B., 1951, *Proc. Instn. Elect. Engrs.*, **98,** Pt. IV, Monograph 2.
Sorge, J., 1924, *Ark. Elektrotech.*, **13,** 189.

LETTERS TO THE EDITOR

Some Properties of Silicon Point-Contact Transistors

A serious disadvantage of the germanium transistor is the rapid deterioration of its performance characteristics with rise in temperature. The forbidden energy gap width of silicon is considerably greater than that of germanium and theoretical considerations show that transistors made with suitable silicon crystals should operate satisfactorily at temperatures higher than those at which a germanium transistor normally fails, i.e. 60–70°c. This has now been confirmed experimentally with silicon point-contact transistors using high back voltage p-type silicon of resistivity 140 ohm cm. (The silicon specimen was supplied by the General Electric Company, Wembley.)

The best transistor action was obtained after etching the silicon surface with a CP-4 mixture (Haynes and Shockley 1951). Phosphor bronze or tungsten probes with a separation of some 0·002 in. were employed and neither the emitter nor collector contacts were subjected to any forming processes. The performance characteristics (Shockley 1950) of a typical transistor made with this high resistivity silicon are as follows: maximum power gain available 8 maximum voltage gain 10, current gain 1·8. However, transistors have been obtained whose performance figures were considerably better than those quoted. Preliminary investigations show that the frequency response does not differ appreciably from that typical of the germanium transistor.

The behaviour of the present silicon transistors has been studied as a function of temperature. The voltage gain at a given operating point decreases to 50% of its room temperature value at about 90°c and to 20% at 150°c. The current gain remains substantially unaltered over this temperature range, the cycle being reversible. Thus, transistor action still obtains in these specimens at elevated temperatures.

The performance at room temperature of the present silicon transistors is clearly inferior to that of the best germanium point contact transistors. However, their ability to operate at high ambient temperatures should prove of practical importance in many circuit applications.

A fuller report of this work will appear elsewhere.

The authors are indebted to the Chief Scientist, Ministry of Supply, and to the Controller, H.M. Stationery Office for permission to publish this letter.

Telecommunications Research Establishment,
Great Malvern, Worcestershire.
25th February 1953.

J. W. Granville.
W. Bardsley.

Haynes, J. R., and Shockley, W., 1951, *Phys. Rev.*, **81,** 835.
Shockley, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

The Influence of Mobility Variation in High Fields on the Diffusion Theory of Rectifier Barriers

In the usual form of the diffusion theories of Mott and of Schottky for the current flow through potential barriers it is assumed that the drift velocity v of the current carriers is linearly related to the electric field E by a constant factor μ, the mobility. Since the field in most of the barrier layer usually exceeds 10^4 v/cm and is often much higher, it is clearly necessary to take into account the reduction of mobility which can occur in fields of this order.

Specifically for the case of germanium Shockley (1951) has reported that the electron drift velocity for fields above 1000 v/cm ceases to be proportional to the field and for fields greater than 3000 v/cm the drift velocity assumes a constant value independent of E.

It is therefore important to examine the implications of a constant drift velocity in the diffusion equation for a rectifying contact between a metal and n-type germanium. This will be done with the following simplifying assumptions : plane geometry, negligible hole current and the validity of the Einstein relation ($D\boldsymbol{e} = \boldsymbol{k}T\mu$); this latter hypothesis requires experimental investigation in high fields.

The current density from metal to semiconductor is then

$$J = -n\boldsymbol{e}v + D\boldsymbol{e}\frac{dn}{dx} = -n\boldsymbol{e}v + \boldsymbol{k}T\mu\frac{dn}{dx},$$

where n is the electron concentration and the coordinate x is measured from the interface into the semiconductor. The drift velocity v is positive here and is always in the direction of increasing x. Furthermore if ψ is the electrostatic potential for electrons

$$v = \mu E = -\mu\frac{d\psi}{dx} \quad \text{whence} \quad J = -n\boldsymbol{e}v - \boldsymbol{k}Tv\frac{dn}{d\psi}.$$

The diffusion equation is now a linear differential equation of the first order relating n and ψ *but not the spatial coordinate* x. This may be integrated in the usual fashion, taking as the boundary conditions

at $x = 0$, $n = n_0$, $\psi = \psi_0$ at the interface
at $x = b$, $n = n_1$, $\psi = \psi_1$ at the end of the barrier layer

with $n_0/n_1 = \exp(-\boldsymbol{e}\phi/\boldsymbol{k}T)$, where ϕ is the effective barrier height and $\psi_0 - \psi_1 = \phi - V$; V is the applied voltage, taken as positive when the metal is positive (forward direction).

Within the framework of the initial assumptions the current density is found, without approximation, to have the form associated with a Mott layer :

$$J = n_0\boldsymbol{e}v\frac{\exp(\boldsymbol{e}V/\boldsymbol{k}T) - 1}{1 - \exp\{\boldsymbol{e}(V-\phi)/\boldsymbol{k}T\}}$$

but this involves no specific assumption about the distribution of ions in the barrier, i.e. about the spatial variation of the potential ψ.

For reverse voltages ($V < 0$) the denominator may be taken as unity and the current density assumes the usual ideal form. As the reverse voltage is increased the saturation current density ($-n_0\boldsymbol{e}v$) increases due to the enhancement of n_0 arising from lowering of the barrier by image force, from the Zener effect and from heating, and also due to electrons tunnelling through the barrier.

Since the limiting drift velocity v is for germanium not greatly different from the thermal velocity of the electrons there is in this case no significant difference in the predictions of the diffusion and diode theories of rectification.

The work described above was carried out as part of the programme of the Radio Research Board, and this note is published by permission of the Department of Scientific and Industrial Research.

Radio Research Station,
Slough, Bucks.
10th February 1953.
R. E. Burgess.

Shockley, W., 1951, *Bell Syst. Tech. J.*, **30**, 990.

A Note on Secondary Processes in a Low-Frequency Electrodeless Discharge

The mechanism for the production of pulses in a low-frequency electrodeless discharge has been well developed recently by Harries and von Engel (1951) and others (Deb and Ghosh 1948). Detailed studies of the nature and occurrence of these pulses in iodine vapour discharge at various potentials elucidated the various secondary processes responsible for the maintenance of discharge under investigation, on which no information existed in the literature. Iodine vapour was excited in cylindrical glass vessels, by 50 c/s potentials fed to external sleeve electrodes; the current structure was investigated by a cathode-ray oscillograph. When the applied voltage V was just equal to the voltage, V_m, sufficient to cause a self-maintained discharge, pulses which were longer than, and distinct from, those to be reported below were noticed on an otherwise smooth sinusoidal current trace (fig. 1 (*a*)). These longer pulses (i) were attributed (Saxena and Ramaiah 1952) to avalanches created by electrons released from the (instantaneous) cathode by well-known γ and $\eta\theta g$ mechanisms (Loeb 1936, 1939). Increase of V above V_m decreased the number of (i) and initiated a different kind of shorter pulses (ii) due to avalanches produced by electrons generated by other β mechanisms occurring in the gas phase (fig. 1 (*b*)). External irradiation initiated (at $V<V_m$, in the parasitic or Geiger region (Korff 1948)) or increased the number of shorter pulses (ii), while inhibiting the longer pulses (Saxena and Ramaiah 1952). When V was increased much beyond V_m (fig. 1 (*c*)), only one of the longer pulses (which persisted over a wide range of V) could be observed at the beginning of that phase of the alternating potential causing ionization by collision; almost the entire current trace was superimposed by (ii). Under these conditions, the authors detected, in the latter part of the half cycle (fig. 1 (*c*)), a third type of pulse (iii) of roughly the same height as (i); this, as far as the authors are aware, has not been reported before. It may be mentioned that (i) and (ii) occurred regularly one after another, while pulses of type (iii) were interposed among (ii). Further, external irradiation had no effect on (iii); the origin of these pulses (iii) needs further consideration.

The favourable influence of enhanced V on the production of the shorter pulses (ii) indicated that, in contrast to a d.c. discharge in non-heterogeneous gases (Loeb 1936), photo-ionization in the gas phase was a significant mechanism for the maintenance of a low-frequency electrodeless discharge, for the inception of which, it would appear, the γ and $\eta\theta g$ processes played an important role. That is, the photons produced in the *primary* electron avalanches created by these last mechanisms are absorbed by gas particles to yield electrons which give rise to further avalanches. This photo-ionization in the gas phase appeared to be inappreciable at low V near V_m and marked at enhanced V. The inhibitive action of large fields on the longer pulses or on the mechanisms responsible for them is a characteristic of the present type of discharge, and follows from a suggestion due to one of us (Ramaiah 1951) that enhanced ionic bombardment on the electrode walls causes dissociation of surface adsorbed molecules to yield atoms and/or radicals which capture the 'secondary' electrons [contributed by those deposited in the preceding half cycle and remaining unneutralized on account of the dielectric nature of glass (Ramaiah and Tiwari 1952)] to be released by γ and $\eta\theta g$ processes and responsible for (i).

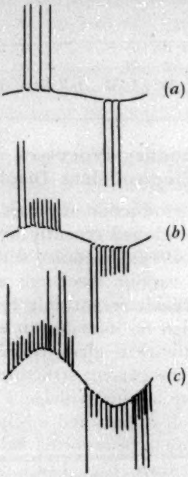


Fig. 1. Oscillograms representing the current structure of a low-frequency electrodeless discharge in iodine vapour at (*a*) 0·9, (*b*) 2·4 and (*c*) 7·2 kv (50 c/s).

Physics Department,
Vaish College, Bhivani.
Faculty of Science,
University of Delhi, India.
27th December 1952.

O. P. DOGRA.
M. G. BHATAWDEKAR.
N. A. RAMAIAH.

DEB, S., and GHOSH, N., 1948, *J. Indian Chem. Soc.*, **25**, 449.
HARRIES, W. L., and VON ENGEL, A., 1951, *Proc. Phys. Soc.* B, **64**, 916.
KORFF, S. A., 1948, *Electron and Nuclear Counters* (New York : Van Nostrand).
LOEB ,L. B., 1936, *Rev. Mod. Phys.*, **8**, 267; 1939, *Fundamental Processes in Electrical Discharge in Gases* (New York : John Wiley).
RAMAIAH, N. A., 1951, *J. Sci. Industr. Res.* A, **10**, 182.
RAMAIAH, N. A., and TIWARI, G. S., 1952, *J. Sci. Industr. Res.* B, **11**, 392.
SAXENA, A. P., and RAMAIAH, N. A., 1952, *J. Chem. Phys.*, **20**, 1342.

REVIEWS OF BOOKS

The Scientific Work of René Descartes (1596–1650), by J. F. SCOTT. Pp. 211. (London: Taylor and Francis, 1952). 20*s.*

Descartes is a significant figure in the history of science because he was one of the most gifted and influential of those remarkable men who began the process by which science became separated from other kinds of philosophy. In his day the scientist was essentially a philosopher, and Descartes recognized no division between what we now call his physics on the one hand and his metaphysics on the other. The same might be said of his great contemporary Galileo, yet today Galileo appears only in histories of science and Descartes only in histories of philosophy. It will therefore probably be a surprise to many that so much of Descartes' attention was given to such subjects as optics, meteorology, physiology, astronomy and mechanics, to say nothing of mathematics in which his contributions to the foundations of coordinate geometry are outstanding.

The explanation of Descartes' sterility in most of these subjects seems to lie in the view which he held of the functions of reason and observation in philosophy. Both Descartes and Galileo were profoundly impressed by the importance of pure reason, especially in its most minutely developed form—that of pure mathematics—but whereas Galileo used his reason to co-ordinate sense observations, and especially those obtained under controlled conditions, or *experimental* observations, Descartes took a poor view of the senses and trusted to reason to tell him what they could apprehend only imperfectly. As always with such efforts, his conclusions were usually far from the truth. His theory of vortices, for example, by which he sought to explain the motions of the heavenly bodies, though for some time a serious rival to the Newtonian theory of gravitation, is now a mere curiosity.

Nevertheless, Descartes has some successes to his credit. Whether he or Snell is to be given chief honour for the discovery of the law of refraction of light may be doubtful, but there is no question that Descartes' demonstration of the law was the more profound and showed the greater insight into the principles involved. He gave the most complete demonstration possible of the formation of the rainbow apart from its colours, and he was a pioneer in meteorology. Though he did not originate the subject of coordinate geometry, his contributions were so vital that he might almost be regarded as its founder. And, perhaps most important of all, though of very questionable advantage, his sharp division between extended things and thinking things—or matter and mind—established the framework in which, even up to the present time, scientists have almost automatically placed their problems before beginning to examine them.

All this, and much more, Dr. Scott brings out admirably in the book before us. No such thorough and trustworthy account of Descartes' scientific work has previously been available. In each subject a short review of earlier work is given, so that Descartes' contributions appear in their true historical setting. Descartes' philosophical system is also outlined; this is essential to an understanding of his work, since he approached all particular problems in the light of his fundamental metaphysical ideas. We obtain, too, some distinct glimpses of Descartes as a man, which do not reveal him as an altogether admirable character. He recognized little merit in his contemporaries, and was not only eager to disclaim any indebtedness to them but also took pains to belittle their achievements. Since they included men like Galileo, Fermat and Pascal, we have now no difficulty in assessing the value of his disparagement. Nevertheless, Dr. Scott regards his failings as " trivial by the side of his great achievements, or the benefits he conferred upon science ". The reader will be able to form his own opinion on this matter. HERBERT DINGLE.

Electronic and Ionic Impact Phenomena, by H. S. W. MASSEY and E. H. S. BURHOP. Pp. xviii + 669. (Oxford : Clarendon Press, 1952.) 70*s*.

The subject matter of this book lies at the root of our understanding of electrical phenomena in gases. The varied aspects of ionization and electrical discharges are only explicable in terms of the dynamical theory of the motions of electrons and ions in gases, and the study of discharge phenomena has been, and is, an important means of investigating elementary electronic and ionic collisional processes. In turn, knowledge thus gained has been of the greatest importance in the elucidation of problems of the extra-nuclear structure of the atom : this is particularly the case in regard to spectroscopy. Consequently, a modern book which gives a comprehensive account of most aspects of impact phenomena for particles of comparatively low energies is bound to be welcome. It can be said at once that *Electronic and Ionic Impact Phenomena* by Professor H. S. W. Massey and Dr. E. H. S. Burhop has succeeded most excellently in doing this, and this work is likely to be the standard reference book on the subject for some time to come.

The mathematical methods and the general theory of collisions has been treated previously in *The Theory of Atomic Collisions* by Mott and Massey (a second edition of which, with cross references to the present book, has appeared), and the present book deals with the subject more from the experimental side, but wherever possible the experimental data are discussed and interpreted in terms of the modern theory of collisions. It is not possible within the confines of a brief review to deal in detail with such a comprehensive volume of nearly 700 pages, but the scope of the treatment can be seen from the chapter headings : The passage of electrons through gases, cross section ; The experimental analysis of the cross section of the impact of electrons with atoms ; Electron collisions with atoms—theoretical description (this is a particularly helpful and useful chapter for the experimental physicist); Collisions of electrons with molecules ; Reflection and secondary emission from surfaces due to electron bombardment ; Electron collisions involving emission of radiation ; Collisions under gas kinetic conditions ; The passage of homogeneous electrons through gases ; The collision of positive and neutral atoms with surfaces ; Recombination. The preparation of a work of this magnitude, involving as it does a highly detailed survey of a vast number of published papers, must have required considerable time, but by the inclusion of a short appendix with notes the authors have been able to include the most recent work on the decay of electron density and on the nature of the various ions in discharges.

The relevant experiments are considered in some detail, and the actual results of the various authors are given rather than their conclusions. This procedure is an admirable feature of the book, greatly enhancing its value for reference purposes. For example, in discussing the question of ionization of gas atoms in collisions with positive ions, the enormous amount of published experimental work on this particular aspect is reviewed in some detail in Chapter 8, where all the elementary data on cross sections for ionization are given. Such an analysis would perhaps be even more valuable to the reader if set against the background of a critical assessment of the various experimental techniques employed, as there is considerable variation in the published values of the cross sections found by different observers ; this is all the more important today in view of the work of Horton and Millest in 1946. While on the subject of positive ions, it may perhaps be worth pointing out that in the account on page 547 of the calculation of the secondary emission coefficient γ for various metals from the Townsend expression for the growth of an ionization current, or from the resulting breakdown criterion, it should have been noted that the procedure is only valid provided that some test does show that the electron emission was actually due only to impact of positive ions ; this is necessary because a precisely similar expression for the growth of ionization current (and therefore for the static breakdown criterion) is obtained if the emission were due to the incidence of photons, but a knowledge of the shape of the $(\gamma, X/p)$ curve is usually sufficient to establish which of the processes in fact predominated.

Considering the comprehensiveness of the work, the number of misprints in this book is extraordinarily small, but it is perhaps worth pointing out one because this is not immediately obvious from the text. In Appendix 1 on page 649 the sentence saying that Biondi measured the recombination coefficient of electrons containing 1% of argon states that he found the coefficient to be less than 10^{-3} of that observed in pure helium ; this should read *pure argon.* The indexes and references are, however, insufficiently detailed for so valuable a book. The authors deserve every congratulation for undertaking the preparation of an up-to-date account of work in this field and for succeeding so admirably in producing a book which is indispensable to those interested in fundamental collisional processes of electrical phenomena in gases. The general production of the book by the Oxford University Press is excellent.

F. LLEWELLYN JONES.

Numerical Analysis, by D. R. HARTREE. Pp. xiv+287. (Oxford : University Press, 1952.) 30*s*.

Most physics research laboratories are now equipped with some type of desk calculating machine. Unfortunately many physicists have never received any instruction in their use and they either avoid using the machines altogether or else fail to use them in the most efficient manner. One reason for this failure to exploit the full possibilities of numerical calculations has been the lack of any single book which indicated clearly the steps required to carry through the work. Professor Hartree's book satisfies this need very successfully. It provides a practical guide to computing the solutions of the most frequently arising types of numerical problem, ranging from the solution of quadratic equations to the inversion of matrices.

The early chapters explain simply the characteristics of the main types of desk machine and the most efficient ways of using them for the basic arithmetical operations. The subject of finite differences is then treated in a concise but satisfactory manner by making use of the method of finite difference operators. This leads to the derivation of the various formulae of interpolation and numerical integration, which are stated in such a manner that they can be referred to without the necessity of following their derivation. The following chapters deal with the solution of simultaneous linear equations, non-linear algebraic equations and the solution of both ordinary and partial differential equations. Finally there is an introductory account to the preparation of calculations for large electronic calculating machines. This last chapter seems rather out of place; all the rest of the book is designed to be of immediate practical value, whereas this one is only a general descriptive account. The author considers that the analysis of observations and statistics are not part of numerical analysis and does not discuss these topics, except for short sections on smoothing and harmonic analysis.

One of the most valuable features of this book is the large number of examples which are set out and worked in detail. Great attention is paid to questions of accuracy and the location of errors, again with many illustrations. There is also a set of examples on which the reader may practise. The author does not attempt to give all the special techniques which are used by professional computers. Throughout he concentrates on the more commonly used methods and is particularly helpful in showing under what circumstances any one method is to be preferred. The book will provide an excellent basis for a short course in computing for students of mathematics or physics, in addition to providing research workers with a useful reference book. J. C. B.

Advanced Strength of Materials, by J. P. Den Hartog. Pp. ix+379. (New York: McGraw-Hill, 1952.) $8.50, 72*s*. 6*d*.

According to the preface, the purpose of this book is to bridge the gap between elementary textbooks on strength of materials and advanced texts on the theory of elasticity. The approach and content of the book may be gathered from the chapter headings: torsion; rotating discs; membrane stresses in shells; bending of flat plates; beams on elastic foundation; two-dimensional theory of elasticity; the energy method; buckling; miscellaneous topics. It concludes with a set of 217 problems.

Two features of the book deserve mention. In the first place, difficulties are firmly and squarely faced and there is no attempt to deceive the reader by unsound argument. Approximations and simplifications are always pin-pointed and their effect on the final result is discussed wherever possible. Secondly, the author is free from inhibitions as far as his style is concerned and two of his wisecracks may be quoted by way of illustration. In connection with the expression for the circumferential strain in a rotating disc, he mentions that the equation " refers to the feelings of a middle-aged gentleman who lets out one notch of his belt after his daily good dinner ". The second quotation may be useful to physicists and engineers when stressed beyond the limit of endurance in Inter-Faculty meetings; it occurs as a comment on a catalogue of results for the bending of flat plates : " In concluding this catalogue we mention an interesting reciprocal theorem. It is contended by some enthusiastic proponents of classical education that if a person has had a good training in Latin and Greek, he is then ready to tackle anything else, such as the theory of flat plates. The reciprocal of this point of view is that if a student has mastered the use of these 25 plate formulae, he has incidentally learned the Greek alphabet and hence is quite ready to start reading and enjoying Attic poetry."

Nowadays, the pressure on university curricula makes it almost impossible to give an adequate training in the theory of elasticity. This book can be recommended without reserve to physicists who, from choice or from force of circumstances, wish to extend their knowledge of statical elasticity and it would serve as a useful and interesting preliminary to the study of the standard texts of Love, Prescott, Southwell and Timoshenko and Goodier on the mathematical theory of elasticity. R. M. D.

Anfangswertprobleme bei partieller Differentialgleichungen, by R. SAUER. Pp. xiii + 239. (Berlin: Springer-Verlag, 1952.) DM. 29.

For information on the mathematical theory of the partial differential equations they encounter, most physicists consult Courant's *Methoden der mathematischen Physik—II*. When this appeared in the thirties much of the extensive material in it on hyperbolic problems, and especially on non-linear ones, may have seemed unnecessary to working physicists. Since then, however, the spectacular expansion of the science of high speed gas flow has provided a wide field of application of this material, and its systematic exposition by Courant certainly facilitated the progress in that field. The same mathematics has been used in modern developments in the theories of plastic flow and of hydraulics.

Now Professor Sauer of München, who has himself made valuable contributions to gas dynamics, has brought out a short self-contained account of the mathematical theory of hyperbolic problems, with detailed specialization of each aspect of the theory for the different branches of physics mentioned above. The book is in four chapters, respectively on (i) classification of problems into elliptic and hyperbolic, (ii) first order partial differential equations, (iii) higher order hyperbolic systems with two independent variables, (iv) systems with three or more. Broadly speaking, the same mathematical material as in Courant's book is covered. There is somewhat more information on the finite difference approach to calculating solutions, using a characteristic network, but no details of computational technique. Little of the more recent work by the Courant school on existence theorems for non-linear hyperbolic problems is included. The Hadamard theory for the linear equation is not abandoned in favour of the streamlined approach due to Marcel Riesz; nor is the theory of ' distributions ' used much to simplify Hadamard's arguments, not at least as much as the preface would lead one to expect.

But the student will find the book pleasantly concise, systematic and easy to follow, and all workers in fields where the material is relevant should find it convenient to use as a work of reference. M. J. LIGHTHILL.

Molecular Theory of Fluids, by H. S. GREEN. Pp. viii + 264. (Amsterdam: North-Holland Publishing Co., 1952.) 20*s*.

This volume is one of a series of monographs (entitled 'Deformation and Flow') on the *Rheological Behaviour of Natural and Synthetic Products*, edited by J. M. Burgers (Delft), J. J. Hermans (Groningen) and G. W. Scott Blair (Reading). The author is professor of mathematical physics at the University of Adelaide, and is probably best known for his work with Professor Max Born on the kinetic theory of liquids.

The book gives an authoritative summary of the present state of the theory of fluids (which includes gases, liquids and non-crystalline solids), both at rest and in motion, both simple and complex, including the theory of diffusion. The final chapters deal with the kinetic theory and the quantum theory of fluids, the bearing of the latter on fluids at very low temperatures and the remarkable flow properties of liquid helium. The preceding chapter discusses elasticity, surface tension, the Brownian motion, the second virial coefficient, gravity and boundary forces and the dielectric constant.

A rigorous quantitative treatment of this wide field necessarily involves much mathematics, which is concisely presented, but nearly a third of the book is non-mathematical text, in which the essential features of the physics of the subject are explained in a very clear and attractive way. The book is a highly valuable addition to the literature of physics. S. CHAPMAN.

Advanced Experiments in Practical Physics, 2nd edn., by J. E. CALTHROP. Pp. xvi + 142. (London: Heinemann, 1952.) 10*s*. 6*d*.

The first edition of this book was reviewed in 1939 (*Proc. Phys. Soc.*, **51,** 894). The main alterations in the present edition are in the section on Electricity and Magnetism, where some experiments on the use of a resonance circuit are described.

This book should be useful to a lecturer or demonstrator in charge of a practical class designed for students who have just taken the intermediate degree examination. He will find therein brief but useful descriptions of a considerable number of suitable experiments, and while few will wish to accept the course as a whole, many will find something to use. All students at this stage would gain by reading the introductory pages entitled 'Instructions to Students', though few students will accept the advice given until they have experienced the results of ignoring it!

The reviewer would prefer to see a larger proportion of experiments on accurate electrical measurements. The buzzer excited resonance circuit has long been obsolete. Some indication should be given as to which of the different methods described for finding the cardinal points of a lens system is intended for a given class of system.

This (and some other books on practical physics) would be improved by a few carefully thought out paragraphs on safety precautions instead of occasional remarks on this matter, which are not very helpful.

Not the least of the merits of this useful book is its low price.

R. W. DITCHBURN.

Filter Design Data for Communication Engineers, by J. H. MOLE. Pp. xvi + 252, 127 figures. (London: Spon, 1952.) 63*s*.

As the title implies, this book is a collection of tables and graphs interspersed with short explanations. The only counts upon which it may be reviewed are the contents, accuracy and layout of the information presented.

Apart from Chapter 12, the filters dealt with are of the Zobel type, and for this reason little assistance is offered to the designer of filters for which insertion loss methods may be required. Within the framework of Zobel type filters it is difficult to imagine anyone requiring any more than is presented here. It is felt however that in the interests of completeness and compactness, the author has often had to restrict the amount of explanations accompanying diagrams. The attempt by the author to achieve completeness (inevitably at the expense of ease of use) puts the work into the category of highly specialized literature, written, in this case, for the full time filter designer, who alone will find it profitable to master the terminology used.

With respect to the accuracy of the tables and graphs which number 56 and 86 respectively, the reviewer's task is an impossible one, and no attempt at checking of figures has been made. Potential users will often have to take the figures on trust as in most cases it is not obvious what method was used in their calculation.

The presentation and pagination are not good; for example much labour is required to discover a method of using the information presented in figures 3 and 4. One feels that the tabular and graphical information would have a greater utility if accompanied by a little more written information.

There is an annoying feature of this book, namely that when referred to in the text, the page on which a figure or table is to be found is not stated. In a book of this kind this can cause a considerable waste of time.

Much of the information presented appears in book form for the first time, and for this reason the book fills a long standing gap in the literature of Zobel type filters.

Chapters 1 to 3 deal with the design of image parameter sections, Chapter 4 with impedance transformations. Chapters 5 to 7 are concerned with the calculation and reduction of reflection losses at the terminals of filters. Chapter 8 is devoted to a study of many section filters particularly those with Tchebycheff stop-band behaviour. The effects of dissipation and component tolerance are discussed in Chapters 9 and 10 respectively which are probably the most valuable sections of this book together with Chapter 11 giving tables of functions, useful to the filter designer. The concluding chapter discusses single sections designed by insertion loss methods.

The binding and paper used are such that they will remain in good condition even with the rough handling to which reference books are subjected.

J. H. WENSLEY.

Fatigue and Fracture of Metals, edited by W. M. MURRAY. Pp. viii + 313. (Massachusetts: Technology Press; New York: John Wiley; London: Chapman and Hall, 1952.) 48*s*.

This book is a collection of fourteen papers read at a conference at the Massachusetts Institute of Technology in June 1950. They include review articles dealing with fatigue and brittle fracture in aircraft, ships, and machinery, with techniques and with the statistical analysis of results. There are also accounts of original researches on various aspects of the problem. A provocative contribution by P. L. Teed—and indeed many of the other papers also—draws attention to the immense gap that still exists between the most complex problem which can be fully treated by a metal physicist, and the simplest situation in engineering service. Physicists will probably find most interesting the paper by E. Orowan on 'Fundamentals of Brittle Behaviour in Metals' which demonstrates with characteristic simplicity and lucidity what can be done to bridge that gap.

N. T.

CONTENTS OF SECTION A

PAGE

Mr. R. H. Tredgold. The Application of Kirkwood's Approximation to the Calculation of Intrinsic Magnetization 421

Dr. L. R. B. Elton and Mr. K. Parker. The Scattering of Fast Positrons by Nuclei 428

Mr. H. C. Rowlinson and Dr. R. F. Barrow. The Band-Spectrum of Aluminium Monofluoride 437

Mr. F. R. Barclay and Mr. W. J. Whitehouse. The Average Number of Neutrons Emitted in the Spontaneous Fission of ^{242}Cm 447

Dr. J. V. Jelley and Mr. W. J. Whitehouse. The Time Distribution of Delayed Particles in Extensive Air Showers using a Liquid Scintillation Counter of Large Area 454

Dr. J. R. Holt and Mr. T. N. Marsham. An Investigation of (d, p) Stripping Reactions—III : Results for ^{28}Si and ^{32}S 467

Mr. H. C. Newns. Polarization Effects in (d, p) and (d, n) Reactions . . 477

Mr. W. A. Runciman and Mr. E. G. Steward. The Atomic Position and Size of the Thallium Ions in KCl(Tl) Phosphors 484

Research Notes :

Dr. A. B. Lidiard. On a Minimum Property of the Free Energy . . 492

Letters to the Editor :

Dr. K. F. Chackett, Dr. J. H. Fremlin and Dr. D. Walker. Nuclear Reactions Produced by Fast Nitrogen Ions 495

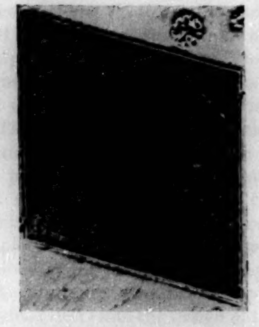
Reviews of Books 496

Contents of Section B 500

PROC. PHYS. SOC. VOL. 66, PT. 5—B (AJIT RAM VERMA AND P. M. REYNOLDS)

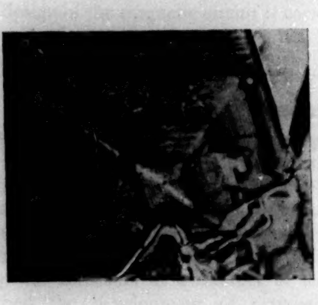
×150

Fig. 1.



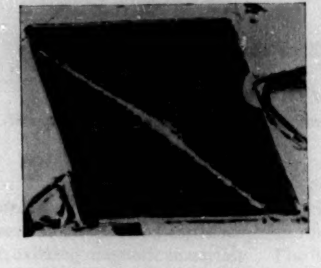
×180

Fig. 2.



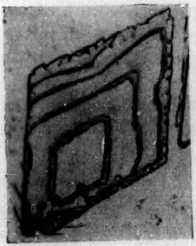
×180

Fig. 4.



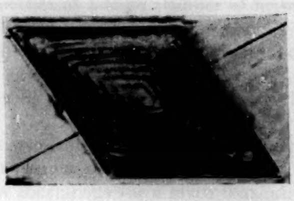
×225

Fig. 5.



×150

Fig. 6.



×360

Fig. 7.

×96

Fig. 8.

The Magnetic Circuit in Electron Microscope Lenses

By T. MULVEY
Research Laboratory, Associated Electrical Industries, Ltd., Aldermaston, Berks.

Communicated by G. Liebmann; MS. received 20th November 1952

Abstract. Previous authors have described the imaging properties of magnetic lenses for the electron microscope and have calculated the best lens performance with existing magnetic materials. The best lens performance can only be obtained if the design of the magnetic circuit is correct. Suitable design curves are given, based upon flux measurements in models. The effect of iron saturation is briefly described.

§ 1. Introduction

The imaging properties of the magnetic lens in an electron microscope depend on the magnitude and distribution of the magnetic field on the lens axis. Glaser, Ramberg and others have treated this problem by evaluating the optical properties of certain analytical field distributions that approximate to the true fields. Dosse (1941) and van Ments and le Poole (1947) have measured the axial distribution of the field experimentally and have evaluated the optical constants of a number of lenses. Ruska (1934, 1942, 1944) has described the development of magnetic lenses for the Siemens electron microscope and has measured the focal properties of a large number of lenses. Recently, Lenz (1950), using the relaxation method, and Liebmann and Grad (1951), using a resistance network, have obtained the axial field distribution with great accuracy, and have computed the imaging properties of a number of magnetic lenses. Liebmann has also evaluated the magnetic lens having the least spherical aberration (Liebmann 1951).

Thus considerable information is now available for the electron-optical design of magnetic lenses. It is possible to predict the best arrangements of pole piece dimensions to give minimum lens aberrations for a given maximum field strength in the air gap corresponding to the maximum flux density permissible in the iron. Attention must be given, however, to the design of the complete iron circuit of the lens to ensure that the desired field strength is obtained in the air gap. This aspect of lens design has received little attention in the literature; the nearest approach to this subject is the investigation by Dreyfus (1931) into the design of the iron circuit for large electromagnets working at very high air gap flux densities.

In the present paper the design of the magnetic circuit will be discussed in relation to magnetic electron lenses of high resolving power and short focal length.

§ 2. Optimum Parameters of Idealized Lenses

In an electron microscope using axially symmetrical lenses the least resolved distance d is set by spherical aberration and diffraction to a value of $d=BC_S^{1/4}\lambda^{3/4}$, where C_S is the spherical aberration constant of the lens and λ the electron wavelength. B is a numerical constant roughly equal to unity: the most recently proposed value is 0·78 (Glaser 1949). Liebmann and Grad (1950) express the optical constants of a lens in terms of the pole piece bore D, the air gap spacing S, and the magnetic field strength H_p at the pole face in the parallel part of the air gap. From these authors' data, the design curves of figs. 1 (*a*), (*b*), (*c*) have been plotted, in order to select a suitable range of operating conditions for the design of the magnetic circuit. The resolution parameter $d/B = C_s^{1/4}\lambda^{3/4}$, at 50 kv, is plotted against the lens bore D for three values of the field strength H_p. The dotted lines indicate the ampere turns needed to excite the air gap to the required field strength. Three sets of curves are shown for S/D values of 0·6, 1·0, and 2·0 respectively. A minimum value of the resolution parameter d/B occurs for each geometry and field strength. The excitation required for minimum d/B departs little from 6000 ampere turns. For an accelerating voltage of 500 kv this excitation is only increased to 12 000 ampere turns as shown in fig. 1 (*d*). Here d/B is plotted against the lens bore D at several accelerating voltages for a constant field strength ($H_p = 24\,000$ oersteds) and for $S/D = 1$.

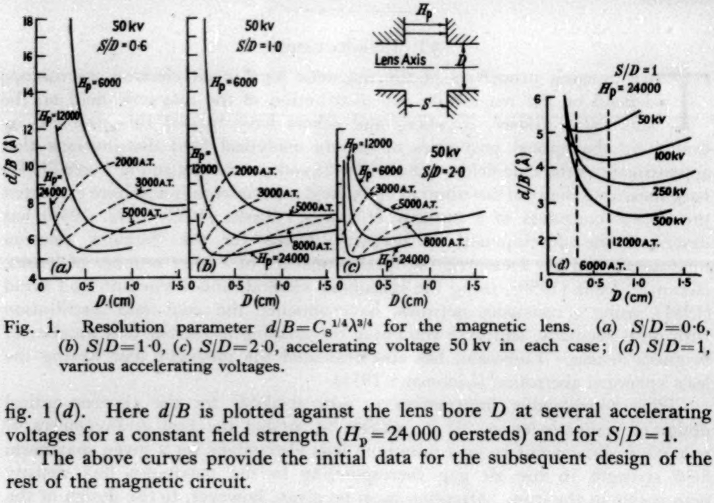


Fig. 1. Resolution parameter $d/B = C_s^{1/4}\lambda^{3/4}$ for the magnetic lens. (*a*) $S/D = 0·6$, (*b*) $S/D = 1·0$, (*c*) $S/D = 2·0$, accelerating voltage 50 kv in each case; (*d*) $S/D = 1$, various accelerating voltages.

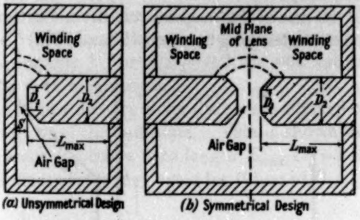
The above curves provide the initial data for the subsequent design of the rest of the magnetic circuit.

§ 3. Representation of Data

Two extreme arrangements of the magnetic circuit may be employed in magnetic lenses for the electron microscope. The first has the air gap at one end of the lens and will be referred to as the *unsymmetrical** design (fig. 2 (*a*)). It is frequently adopted in objective lenses in order to simplify the mechanical stage construction. The second, which is often encountered in projector lenses, has the air gap at the centre of the lens and will be referred to as the *symmetrical* design (fig. 2 (*b*)). Here, owing to the mirror symmetry of the magnetic field about the mid-plane of the lens, only one half of the lens need be considered. Further, since the mid-plane of the lens is an equipotential surface, it may be replaced by a plane of infinite permeability. It is therefore only necessary to provide data for the unsymmetrical design. The design curves in this paper consequently refer to *the unsymmetrical design in the absence of the lens bore*; the effect of the bore on the curves is considered separately. It may be noted here that when the dimensions and the ampere turns of a magnetic lens are scaled by a factor N, the field at corresponding points in the scaled and original model remains unchanged. Lens dimensions are given in relative units in terms of the pole face diameter D_1. The relative flux density B/B_p is defined as the ratio between the flux density B at any point in the iron and the flux density B_p in the parallel part of the air gap. B_p is given by $B_p = 4\pi NI/10S$, where NI is the number of ampere turns in the air gap.

---

* The symmetry of the complete iron circuit should not be confused with the symmetry of the axial field distribution.



(*a*) Unsymmetrical Design (*b*) Symmetrical Design

Fig. 2. Magnetic circuit arrangements.

§4. Method of Investigation

Flux measurements were made on experimental lens models in which all parts of the magnetic circuit were accessible, and the shapes of which could be altered. Most measurements were made under unsaturated conditions in the iron. Separate measurements were made under conditions of pole piece saturation.

In order to measure the flux distribution, search coils were placed at various points in the magnetic circuit and the flux traversing them was measured with a fluxmeter.

§5. The Magnetic Circuit

Figure 3 (*a*) shows a simple unsymmetrical magnetic circuit that illustrates several important aspects of magnetic design. An air gap of length S is formed between a flat end-plate and solid cylindrical central core of diameter D_1 and length L_{max}. The return path for the flux is not shown. The flux entering the pole face near the axis is approximately parallel. Leakage flux from the end-plate enters the curved sides of the cylinder as shown schematically in fig. 3 (*a*). Figure 3 (*b*) shows how the relative flux density B/B_p through any cross section varies with the relative axial distance L/D_1 from the pole face for three values of the ratio of gap length to pole face diameter S/D_1. The total flux, and therefore the flux density in the cylinder, increases with increasing values of L/D_1. Consequently in such a design premature saturation of the iron occurs in the part of the core farthest from the air gap. This sets a limit to the maximum air gap flux density, which in extreme cases may be a small fraction of the maximum flux density in the iron.

Where high flux densities are required in the air gap, the flux density in the iron must be kept below that in the air gap. This may be achieved by tapering the pole piece, as shown in fig. 4 (*a*), so as to increase the iron cross section. Figure (4 *b*) shows the relative flux density B/B_p at a distance L/D_1 along the pole piece, for taper-angles of 25°, 45°, 55° and 70° respectively, with $S/D_1 = 0{\cdot}2$. When the angle θ is small, e.g. 25°, the increase in iron cross section is not sufficient to prevent the initial rise in flux density. On the other hand, if θ is made greater than 70° the total flux intercepted by the pole piece becomes large, necessitating correspondingly larger dimensions for the rest of the magnetic circuit. Further discussion, therefore, will be restricted to angles lying between 45° and 70°.

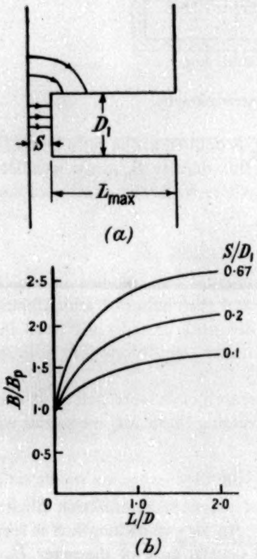


Fig. 3. (*a*) Simple magnetic circuit, (*b*) flux density distribution along the central core of a simple magnetic circuit.

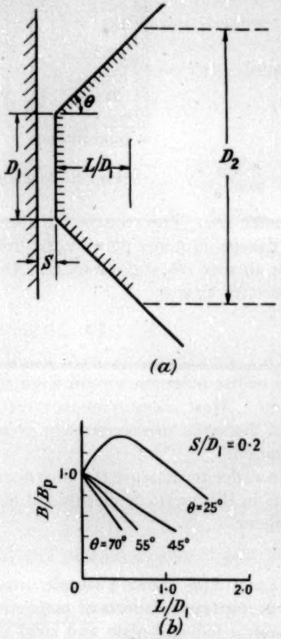
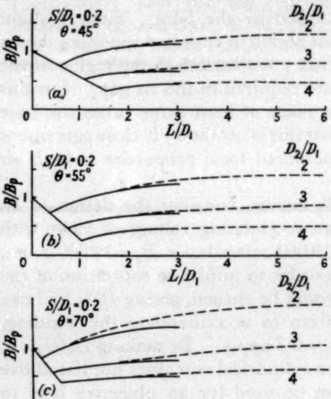


Fig. 4. (*a*) Pole piece with taper-angle θ, (*b*) effect of taper-angle θ on the flux density distribution along the pole piece.

When the relative flux density has fallen to a small enough value, the taper may be stopped and the core continued as a cylinder of diameter D_2, as shown by the dotted lines in fig. 4 (*a*). The flux density will now increase along the cylindrical part of the core. The resulting distribution along the core is shown in fig. 5 (*a*) for $S/D_1 = 0{\cdot}2$ and $\theta = 45°$. Three curves are shown for values of D_2/D_1 of 2, 3 and 4 respectively. The value of D_2/D_1 fixes the outside diameter of the core. Since the power dissipated in the energizing coil increases with increasing values of D_2/D_1 an upper limit is set for this ratio. The smallest value of D_2/D_1 is determined by the greatest flux density that can be allowed at the root of the core. If the flux density at the root of the core, or elsewhere, is allowed to exceed that in the gap, then at high excitations this part of the iron circuit will have a low permeability, resulting in a loss of ampere turns. Saturation of any part of the iron circuit will limit the maximum air gap flux density. A good design will allow a factor of safety by restricting the flux density in the core to about one-half to two-thirds of that in the air gap. For the design illustrated, a value of $D_2/D_1 = 3$ is adequate.

Figures 5 (*b*) and 5 (*c*) show the corresponding distributions for $S/D_1 = 0{\cdot}2$ with $\theta = 55°$ and 70° respectively. For $S/D_1 = 0{\cdot}2$ the safe value of D_2/D_1 is seen to increase only slightly with taper-angle.

The distribution of flux density along the core is affected by the finite length of the core and the presence of the end-plate. This is shown in the curves of fig. 5. The full lines represent a relative core length $L_{\max}/D_1 = 3$ and the broken lines $L_{\max}/D_1 = 6$. The effect may be neglected for values of D_2/D_1 greater than 3.



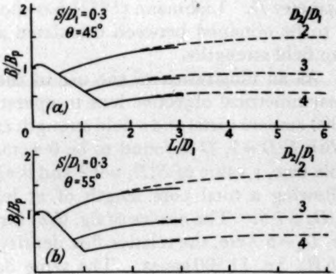


Fig. 5. Flux density distribution along the pole piece and central core with $S/D_1 = 0{\cdot}2$ and for taper-angles θ of 45°, 55° and 70°.

Fig. 6. Flux density distribution along the pole piece and central core for $S/D_1 = 0{\cdot}3$ for taper-angles of 45° and 55°.

Figure 6 shows the effect of increasing the S/D_1 ratio to 0·3, using angles of 45° and 55° respectively. The curves are in general similar to those of fig. 5 but the relative flux density at corresponding points is greater, since the leakage flux becomes greater as the air gap is widened. A slight increase above unity in relative flux density now takes place in the pole piece at small values of L/D_1. For $\theta = 55°$ this increase amounts to a few per cent above the air gap flux density. Thus, to a first approximation, the gap width S/D_1 should not exceed 0·3 if the flux density in the iron circuit is to be everywhere less than that in the air gap.

§ 6. THE EFFECT OF THE LENS BORE

So far, the effect of the bore on the flux density distribution has not been considered. When a hole of diameter D is bored through the pole face, the parallel flux is redistributed and a strong fringing field arises at the edge of the hole. A detailed discussion of the field distribution at this edge, and its effect on the imaging properties, is given by Liebmann (1953). In order to study how the redistribution of flux affects the design of the magnetic circuit, a pole piece model was made with a parallel bore into which an accurately machined iron plunger was fitted. The flat end of the plunger was arranged to be level with the pole face. On withdrawing the plunger, the change in flux distribution in the pole piece could be measured by means of suitably placed measuring coils. It was found that in the range of S/D_1 between 0·1 and 1·0 the reduction in flux entering the pole piece, caused by the bore, was small compared with the resulting reduction in iron cross section. For example, a reduction in iron cross section at the pole face of 33%, caused by the bore, decreased the flux entering the pole tips by only 5%. The main effect of the bore, therefore, is to *increase* the average flux density in the immediate vicinity of the pole face by a fraction approximately equal to the fractional reduction in iron cross section caused by the bore. Since a slight initial rise in flux density may take place in this region in certain designs, as shown previously, it is advisable not to reduce the iron cross section in the region of the pole face too seriously if high flux densities are required in the air gap. For this reason the pole face diameter D_1 should be made at least three times the bore diameter D. Liebmann (1953) has shown that this is necessary if close agreement is to be obtained between calculated and measured focal properties at high air gap field strengths.

As an illustration of the use of the B/B_p curves, consider the design of an unsymmetrical objective lens to operate at an accelerating voltage of 50 kv with 6000 ampere turns and a field strength H_p of 19 000 oersteds, i.e. $B_p = 19\,000$ gauss. With $S/D = 1$, D is found to be 0·4 cm. In order to minimize saturation of the pole face, a value of $S/D_1 = 0{\cdot}3$ and $\theta = 55°$ would be chosen, giving $D_1 = 1{\cdot}33$ cm. Allowing a total core length of at least 10 cm to accommodate the winding, $L/D_1 \geqslant 7{\cdot}5$. The curves of fig. 6 (*b*) for $S/D_1 = 0{\cdot}3$ apply. By making $D_2/D_1 = 4$, i.e. $D_2 = 5{\cdot}3$ cm, the relative flux density in the cylindrical core does not rise above $0{\cdot}6\,B_p$, i.e. 11 500 gauss. The same data can be used for an objective lens to operate at 500 kv with 12 000 ampere turns by scaling the dimensions and the ampere turns by a factor of two.

Consider now the magnetic design of a symmetrical projector lens, to operate up to 150 kv, having a bore of 0·2 cm, an S/D ratio of 1·5 and 4700 ampere turns in the gap.* Here $S = 0{\cdot}3$ cm and $H_p = 19\,700$ oersteds. The design proceeds as for an unsymmetrical lens with $S = 0{\cdot}15$ cm and H_p unchanged. By making $D_1 = 0{\cdot}75$ cm and $\theta = 55°$, the B/B_p curves of fig. 5 (*b*) for $S/D_1 = 0{\cdot}2$ apply. Choosing $D_2/D_1 = 4$, D_2 becomes 3 cm.

§ 7. CONCLUSION

By the correct choice of lens dimensions the magnetic conditions that have been assumed in theoretical calculations can be closely approximated. The magnetic circuit of a lens designed for high resolving power should, in general, satisfy the following conditions. The pole face diameter D_1 should be at least three times as large as the lens bore D. An angle of pole piece taper of between 50° and 60° is usually the best choice. The relative air gap length S/D_1 (referred to the unsymmetrical design) should not exceed 0·3. Probably the most important dimension is the outer diameter D_2 of the central core; this should generally be greater than three times the pole face diameter in order to avoid saturation of the iron circuit.

---

* For the optimum design of magnetic projector lenses see Liebmann (1952).

ACKNOWLEDGMENTS

The author wishes to thank Messrs. P. E. Watts and J. H. Nevitt for their assistance at different times with the flux measurements, and Mr. M. E. Haine and Dr. G. Liebmann for many stimulating discussions. He also wishes to thank the Director of the Laboratory, Dr. T. E. Allibone, for permission to publish this paper.

REFERENCES

DOSSE, J., 1941, *Z. Phys.*, **117,** 437.
DREYFUS, L., 1931, *Arch. Elektrotech.*, **25,** 392.
GLASER, W., 1949, *Acta Phys. Austriaca*, **3,** 38.
LENZ, F., 1950, *Z. angew. Phys.*, **2,** 448.
LIEBMANN, G., 1951, *Proc. Phys. Soc.* B, **64,** 972; 1952, *Ibid.*, **65,** 94; 1953, *Ibid.*, **66,** 448.
LIEBMANN, G., and GRAD, E. M., 1951, *Proc. Phys. Soc.* B, **64,** 956.
VAN MENTS, M., and LE POOLE, J. B., 1947, *Appl. Sci. Res.* B, **1,** 3.
RUSKA, E., 1934, *Z. Phys.*, **89,** 90; 1942, *Arch. Elektrotech.*, **36,** 431; 1944, *Ibid.*, **38,** 102.

The Effect of Pole Piece Saturation in Magnetic Electron Lenses

By G. LIEBMANN

Research Laboratory, Associated Electrical Industries Ltd., Aldermaston, Berks.

MS. received 20th November 1952

Abstract. The imaging properties of magnetic lenses in the absence of iron saturation are now well known, but there has been some doubt up to what flux densities the calculated lens performance would be correct. It is shown here that the effect of iron saturation in the pole piece tips is equivalent to an apparent increase in lens dimensions, apart from a loss of m.m.f. in the iron, which depends on the design of the magnetic circuit as a whole. Published data for unsaturated magnetic lenses can still be used beyond the onset of pole piece saturation if these effects are taken into consideration. A practical limit of lens operation is reached for a flux density in the iron near the air gap which would make $\mu \simeq 50$.

§ 1. Introduction

Several systematic investigations of the imaging properties of symmetrical magnetic electron lenses of given geometry in which infinite permeability of the iron was assumed, so that the imaging properties became unique functions of lens geometry and lens excitation, have been carried out during the last few years (van Ments and le Poole 1947, Lenz 1950 a, Liebmann and Grad 1951, to be referred to as I). The part played by iron saturation in setting a limit to the improvement of such 'ideal' lenses is well known; a theoretical study of a particular magnetic lens under conditions where the pole pieces are highly saturated was carried out by Hesse (1950), and a similar approximation method was used by Lenz (1950 b) to show the 'field broadening' effect of saturation.

In view of the full design data now available for lenses under unsaturated conditions, it was thought desirable to see how far these data can be used before saturation at the pole piece tips makes itself felt, and what quantitative modifications would have to be made in the use of these data under saturation conditions. In the present investigation it has still been assumed that the iron circuit as such is of sufficient cross section everywhere to ensure that the pole piece tips saturate before any other part of the iron circuit. This condition has very often not been satisfied in practice. An investigation of the design of the iron circuit as a whole has been carried out by Mulvey (1953), and the results of this paper are complementary to his work.

§ 2. Method of Investigation

The pole piece assembly of a symmetrical electron lens of gap width S and bore diameter D is shown in fig. 1. The field distribution within such lenses for finite values of permeability was studied by setting up models of the pole pieces (see fig. 2) on a resistance-network analogue; this was of the (r, z) type described by the author (Liebmann 1950), except that wire-wound precision resistors were used in the apparatus, and that the new network was larger. In all models $R = D/2 = 10$

mesh units and $L=100$ mesh units, whereas $S/2$ was varied from 6 to 20 mesh units ($S/D=0{\cdot}6$ to $S/D=2$). The midplane of the lens and the end-plane of the pole piece were assumed to be equipotential surfaces of m.m.f., and a potential difference of $\Psi_0=1$ was applied in the model to these two planes. While this investigation covered different pole piece shapes, the most detailed information was obtained for 'idealized' lenses in which the pole faces are perpendicular to the axis (A – A‴ in fig. 2) up to a great distance from the axis, where the flux lines are parallel to the axis and the flux densities in the air gap and in the iron are identical (numerically equal to H_p of I).

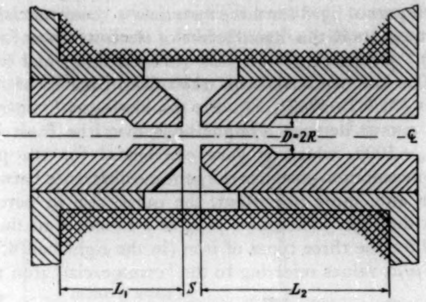


Fig. 1. Pole piece assembly of magnetic electron lens.

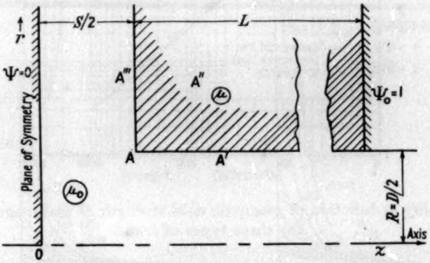


Fig. 2. Pole piece model investigated by resistance-network analogue.

It had been shown (Liebmann 1950) that the resistance-network analogue can solve the equation governing the problem

$$\operatorname{div}(\mu H)=0 \qquad \text{......(1)}$$

if the local values of the network resistances are given the values

$$R_N'=(\mu_0/\mu)\,R_N, \qquad \text{......(2)}$$

where R_N is the value of the network resistance at the point (r, z) representing the permeability of air $\mu=\mu_0$. All those network resistances which occupy the area on the network which corresponds to the space taken up by iron in the lens are therefore shunted to values R_N' as indicated by eqn. (2).

The experimental procedure is to select a flux density B and find the corresponding value of μ/μ_0 from the (B, H) curve of the iron used for the pole pieces and then carry out the shunting of the network resistances as required. As it is awkward to measure flux densities (corresponding to currents) within the resistance network, but easy to measure field strengths as the potential difference between adjacent network studs (taking the third difference correction into account if high accuracy is required), it is more practicable to use the $(\mu/\mu_0, H)$ relationship. This is shown in the left part of fig. 3 for three materials, a ' commercial ' iron which has been used for some years in the manufacture of electron lenses for electron microscopes (full line), ' electrolytic ' soft iron (broken line), and 50/50 cobalt–iron (dotted line). The last two have been taken from page 25 of *Magnetic Circuits and Transformers* (M.I.T. 1943). In the double-logarithmic representation of fig. 3 the $\mu(H)$ curves become straight lines over the most important range $\mu/\mu_0=10$ to $\mu/\mu_0=1000$, which is very convenient for the present purpose. However, this representation obscures the genuine difference between soft iron and 50/50 cobalt–iron and, more important, the quick rise in permeability with a reduction in flux density. Therefore μ/μ_0 as a function of flux density B has also been plotted for the three types of iron (in the right half of fig. 3). In this investigation the μ/μ_0 values referring to the 'commercial' iron were used.

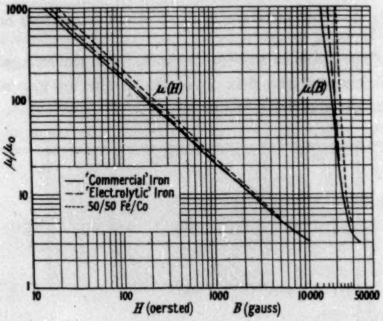


Fig. 3. Permeability as function of magnetic field strength H and magnetic flux density B for three types of iron.

A difficulty arises through the dependence of μ/μ_0 on H. While for the 'idealized' lens the field strength is uniform at a great distance from the air gap, where $H=H_{\mathrm{I}}=(\mu_0/\mu)\,H_{\mathrm{p}}$, it increases strongly near the corner A in fig. 2. Hence different values of μ/μ_0, as obtained from fig. 3, apply in the region A – A′ – A″ – A‴, i.e. different values of R_{N}' would have to be used there, but the field distribution in this region is not known beforehand. One might therefore carry out an approximation process, by first shunting the network for the space taken up by iron, using the *same* value of μ/μ_0 in working out the resistance values R_{N}' from eqn. (2). One would then measure the ensuing field distribution $H(r,z)$ in the network model and carry out modifications of the shunt values, using these measured values of $H(r,z)$ to find the appropriate μ/μ_0 values from the graph, fig. 3. After the resistance modifications a new $H(r,z)$ plot would be obtained, new resistance values worked out, etc. This process would have to be carried on until the field distribution has become 'self-consistent', i.e. until the μ/μ_0 values determined from fig. 3 for the field strength $H(r, z)$ measured in the network are identical with those actually used in working out R_N' for the model from eqn. (2).

In this form the approximation process is not convergent near the corner A. The process is greatly improved by combining it with an extrapolation method. In this, as a first step, uniform permeability is assumed throughout the iron, for example a permeability of $\mu/\mu_0=50$ is assumed, corresponding to $H_I=395$ oersteds. (This corresponds to a field strength in the air gap at great distance from the axis of $H_p=19{\cdot}8$ kilo-oersteds.) Now the *relative* value of the field strength in terms of H_I is measured in the network model, and this ratio $F=F_1$ ('first approximation') is plotted, for example, as a function of z for given values of r = constant (fig. 4 (*a*)). The new values of μ/μ_0 are then read off fig. 3 for each value of F_1, taking into account that $F_1=1$ is equivalent to 395 oersteds. The network resistance shunts are then modified according to these new values of μ/μ_0, and a new plot of relative field strength $F=F_2$ (fig. 4 (*b*)) is taken. No μ/μ_0 values are determined for F_2, but a new distribution of relative field strengths

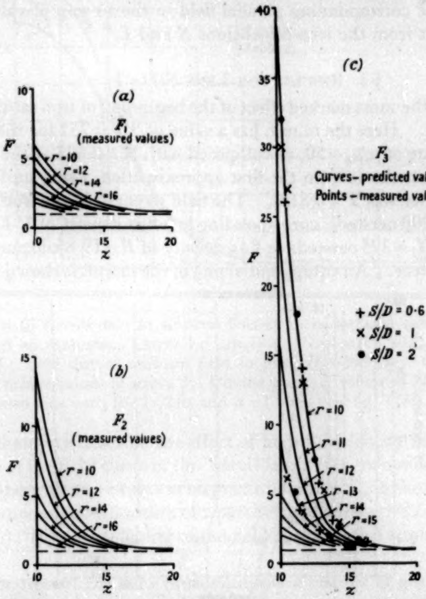


Fig. 4. Field strength increase factor $F=H(r, z)/H_I$ as function of axial distance z for given values of radial distance r at the three stages of the analogue iteration process. (*a*) Initial stage: uniform permeability (measured values of F). (*b*) Second stage: after first analogue modification (measured values of F). (*c*) Third stage: curves give predicted values of F, used in setting up final analogue model; points give values of F measured in this model, and showing 'self-consistency'.

F_3 is calculated from F_1 and F_2 according to the relation

$$F_3/F_2 = F_2/F_1 \,. \qquad \ldots\ldots(3)$$

The second modification of the values of the resistances in the area A – A′ – A″ – A‴ is based on the plot of the F_3 values (fig. 4 (*c*)). Upon measuring the field strength in this modified model it was found that the model was 'self-consistent', as the μ/μ_0 values determined for the measured field strengths were identical within measuring accuracy with the μ/μ_0 values used in working out the modified resistances for the area A – A′ – A″ – A‴. A few typical measured 'final' values of F, bearing out this 'self-consistency', for $S/D = 0{\cdot}6$, 1 and 2, for the same resistance modifications near corner A, are shown in fig. 4 (*c*).

The amount of modification required before 'self-consistency' is reached depends on the initial value of μ/μ_0; it is the greater the smaller the initial μ/μ_0. The excitation, in ampere turns, required to produce the initial value of μ/μ_0, i.e. to produce the corresponding parallel field in the air gap of value H_p, can be easily worked out from the lens dimensions S and L.*

§ 3. RESULTS FOR LENS $S/D = 1$

As expected, the most marked effect of the beginning of iron saturation appears near the corner A. Here the m.m.f. has a value of $\Psi = 0{\cdot}751$ for the lens $S/D = 1$ and an initial value of $\mu/\mu_0 = 50$, as compared with $\Psi = 0{\cdot}833$ at the pole face at a great distance from the axis (in the first approximation, using uniform μ/μ_0, the m.m.f. at the corner was $\Psi = 0{\cdot}812$). The field strength in the iron rises near the corner to $H \simeq 13\,000$ oersteds, corresponding to a flux density of 35 kilogauss, for a field strength of $H_{\rm I} = 395$ oersteds or flux density of $B = 19{\cdot}8$ kilogauss in the main body of the pole piece. An equipotential plot of the m.m.f. is shown in fig. 5.

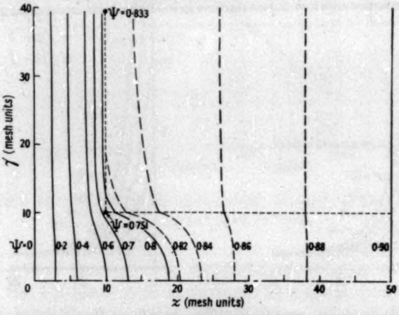


Fig. 5. Equipotential plot of m.m.f. in electron lens at incipient iron saturation $S/D = 1$. $H_{\rm I} = 395$ Oe, $\mu/\mu_0 = 50$.

The lens properties are determined by the distribution of field strength along the axis. For an applied m.m.f. of $\Psi_0 = 1$ the field $(\partial\Psi/\partial z)_0$ along the axis would be represented by curve 1 in fig. 6 for large values of μ/μ_0. An increase in excitation such that $H_{\rm I} \simeq 395$ oersteds at a great distance from the air gap, as in the present example, changes this distribution to that given by curve 2 in fig. 6. The maximum value of $(\partial\Psi/\partial z)_0$ drops slightly more (3·7%) than would correspond to the loss of m.m.f. in the pole pieces for uniform permeability ($\mu/\mu_0=50$), to $(\partial\Psi/\partial z)_{\max}=0{\cdot}669$ instead of to $(H_0/H_p)(\Delta\Psi/\tfrac{1}{2}S)=0{\cdot}834\times0{\cdot}833=0{\cdot}695$ (see fig. 7 of I). In addition, the field curve is slightly broadened, its half width increasing from $a/R_0=1{\cdot}13$ for curve 1 to $a/R_0=1{\cdot}19$ for curve 2, and a 'tail' is added, corresponding to the uniform field H_{I} at a great distance from the air gap, where $\partial\Psi/\partial z=0{\cdot}0167$. Subtracting this parallel field from the measured curve, taking into account the decrease of this parallel field near the gap as obtained by separate measurements and shown in curve 3, one obtains curve 4.

---

* The method given here of obtaining the field distribution within the iron can, of course, be applied to other problems, e.g. the field distribution in the pole pieces of magnets, or in the armatures of electrical machines.

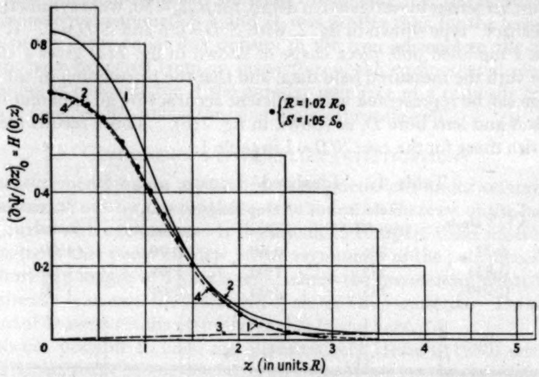


Fig. 6. Axial field distribution in electron lens at incipient iron saturation $S/D=1$, as measured on analogue. Curve 1 : $\mu/\mu_0=\infty$. Curve 2 : $\mu/\mu_0=50$ ($H_{\mathrm{I}}=395$ Oe). Curve 3 : field due to uniform field in iron, $H_{\mathrm{I}}=395$ Oe. Curve 4 : values of curve 2 minus values of curve 3. Circled points : values of $H(z,0)$ for substitute unsaturated lens with $R=1{\cdot}02R_0$ and $S=1{\cdot}05S_0$ (see fig. 7 (*b*)).

One can therefore describe the effect of high flux density in the iron by these modifications of the field curve of the 'ideal' lens: (i) a reduction in the 'effective' number of ampere turns, or value of magnetic potential Ψ_0, in the ratio $1:1+2L/\mu S$, with a simultaneous superposition of a parallel field of length $2L$ and field strength $2\Psi_0/(\mu S+2L)$; (ii) a broadening of the field curve, with a simultaneous decrease in its height.

One can represent the field broadening by a change of geometry of the pole piece tips. Measurements taken on the resistance-network analogue have shown that the changes in the equipotential lines in the air space surrounding point A due to incipient saturation of the pole tips can be represented by a change of the pole faces as shown by the broken lines in fig. 7 (*a*), the outline of the modified pole piece representing the magnetic equipotential line $\Psi=0{\cdot}833$, for $\mu/\mu_0=50$, $L/R=10$. Apart from the superposition of the parallel magnetic field in the lens bore, and the drop of m.m.f. in the pole piece, this modification of the shape of pole piece yields an equipotential plot which is indistinguishable from that obtained for the model representing the correct spatial distribution of μ/μ_0, except quite near the inner surface of the pole piece bore.

From a practical point of view it is more convenient to express the broadening of the field curve and its reduction in height by an increase of the gap length S and a smaller increase of the lens bore $D=2R$, as shown in fig. 7 (*b*), because then it is possible to correlate the lens data for unsaturated and partially saturated pole pieces. In the particular example $\mu/\mu_0=50$ one finds a best representation for the 'idealized' lens by $S=1{\cdot}05\,S_0$, $D=1{\cdot}02\,D_0$, with $\Psi_{\rm pole}=0{\cdot}833$ and $\mu/\mu_0=\infty$, the suffix zero referring to the actual physical dimensions of the lens. The axial field values for this modified pole piece geometry are entered as circled points in fig. 6.

§ 4. OTHER RESULTS

Two further lenses investigated in detail, for $\mu/\mu_0=50$, were symmetrical lenses of the 'idealized' type shown in fig. 2, with $S/D=0{\cdot}6$ and $S/D=2$. It was again found that a modified pole piece shape as shown in fig. 7 (*a*) gives a close correspondence with the measured field data, and that the broadening of the axial field distribution can be represented with sufficient accuracy by an apparent increase in gap length S and lens bore D, as shown in fig. 7 (*b*). These results are collected together with those for the case $S/D=1$ in table 1.

Table 1. 'Idealized' Lenses, $\mu/\mu_0=50$

$(S/D)_0$	$(\Psi/\Psi_0)_{\rm pole}$	Field in iron (Oe)	Gap length ($\times D_0$)		Lens bore ($\times D_0$)	
			actual	apparent	actual	apparent
0·6	0·75	395	0·60	0·70	1·00	1·02
1	0·833	395	1·00	1·05	1·00	1·02
2	0·909	395	2·00	2·04	1·00	1·02

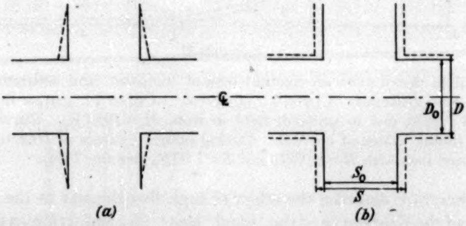


Fig. 7. Equivalent unsaturated pole piece geometries representing saturated pole piece tips. (*a*) Best equivalent geometry. (*b*) Equivalent geometry used in practice (see circled points in fig. 6).

One sees that the apparent increase in lens bore is slight and independent of S/D, whereas the apparent increase in gap length is the more pronounced the shorter the gap.

The three lenses listed in table 1 were also investigated for the condition $\mu/\mu_0=100$, equivalent to a field strength of 185 oersteds in the iron at sufficient distance from the corner. The same effects were found as in the case $\mu/\mu_0=50$, but the changes in field shape etc. are about one third in magnitude; a more important effect in this case may be the drop of m.m.f. in the iron circuit.

Higher initial field strengths than $H=395$ oersteds were not investigated, as for these the modifications required in the resistance-network model would have had to be extended over a much larger part of the network. One would expect to find the same effects as in the case discussed here in detail, except that the field broadening, or equivalent modification of pole piece geometry, and the loss of ampere turns would be greater. However, as seen from the example given in § 6, lenses are not operated in practice under such highly saturated conditions.

A few other lenses where the pole pieces were truncated cones, as in fig. 1, were studied more briefly by the method of § 2. It was again found that the effect of high flux density in the iron could be described by the two modifications given at the end of § 3. If the pole face was left wide enough in the model, i.e. the outer diameter of the face was at least three time the bore diameter, the changes were identical with those for the corresponding 'idealized' lenses of table 1, where then the value of μ/μ_0 refers to the part of the iron adjacent to the pole tip (Mulvey 1953). If the pole face was made smaller, the field broadening, i.e. the increase in the equivalent geometrical parameters S and D, was greater than for the lenses of the type shown in fig. 2 for equal flux density in the iron adjacent to the pole tip. As one might have expected from an inspection of the equipotential plot, fig. 5, field broadening is considerable if the annular pole face of a conically truncated pole piece is of small width, unless the cone angle approaches 90°.

§ 5. Comparison with Earlier Investigations

Field measurements on an asymmetrical magnetic lens under saturation conditions were carried out by Dosse (1941), who found an increase of the half width of the field curve with excitation. It is difficult to compare these measurements with the results of this paper in view of the asymmetry of the pole pieces and the relatively short gap length of Dosse's lens, where the broadening effect becomes noticeable already at an axial field strength of about 7 kilo-oersteds. However, the general trend of Dosse's results is similar to that found here.

It is also not possible to compare quantitatively Hesse's (1950) results with those of this paper, as the geometrical data given by Hesse are not complete, and the case for which detailed field data are given is one of extreme saturation of the pole pieces. The axial field distribution curves for the unsaturated and the saturated pole pieces were therefore plotted from Hesse's data, and from the shapes of these curves relative values of S/D and D were obtained, using the field values for unsaturated symmetrical lenses given in I. From this one finds for Hesse's unsaturated lens, which has a symmetrical field distribution, a value of $S_0/D_0=0{\cdot}54$. Saturation produces considerable asymmetry in Hesse's lens, but the right-hand part of the field distribution can be matched by an equivalent unsaturated field distribution with $S_1=1{\cdot}63\,S_0$ and $D_1=1{\cdot}12\,D_0$. For the left-hand part of the field the equivalent change of geometry is still greater, $S_2\simeq1{\cdot}9\,S_0$ and $D_2\simeq1{\cdot}2\,D_0$. This shows that it is possible to express the modifications of an 'ideal' lens even under conditions of severe iron saturation by the two effects listed in § 3. Another interesting fact is that the ratio air gap change to lens bore change is the same (about 5 : 1 for $S/D\simeq0{\cdot}6$) for incipient saturation as investigated in this paper and for extreme saturation as investigated by Hesse.

The numerical example by which Lenz (1950 b) illustrated his approximation method for the field determination in saturated lenses gives results similar to those of this paper, but again it is difficult to compare the results in detail, mainly owing to the asymmetrical geometry of Lenz's example. If one assumes D equal to the mean value of the two bore diameters of Lenz's lens, the unsaturated condition corresponds to $S/D\simeq0{\cdot}45$, whereas under saturation, corresponding to $\mu/\mu_0\simeq50$, one finds $S/D\simeq0{\cdot}56$. This agrees quite well with the data of our table 1.

§ 6. IMAGING PROPERTIES OF LENSES UNDER INCIPIENT IRON SATURATION

In § 4 it was shown that iron saturation at the pole tips can be described by (i) a reduction in the effective m.m.f. across the air gap ('loss of ampere turns') with the simultaneous appearance of a weak parallel field in the lens bore, and by (ii) an apparent increase in the physical dimensions S and D of the lens. Which of these effects is the more important will depend on the design of the iron circuit. In an 'idealized' lens, with indefinitely large iron cross section, the first of these two effects is by far the stronger at the onset of saturation; the effect of the parallel field in the lens bore is negligible as the refracting power is proportional to $\int H^2 dz$. This is illustrated by the example $S/D=1$, $\mu/\mu_0=50$, with such values of S, D_0, NI and V_r that the lens excitation parameter $k^2=0{\cdot}022\,H_0^2R^2/V_r=1$ for $\mu/\mu_0\to\infty$. From the data given in I and in § 4 of this paper one obtains the data of table 2.

Table 2

No.	μ	f/R_0	z_1/R_0	Remarks
1	∞	1·09	0·68	—
2	50	1·32	0·99	Corrected for drop of m.m.f. across gap
3	50	1·36	1·02	As 2, and corrected for field broadening
4	50	1·35	1·01	As 3, and corrected for parallel field in bore

The data given in the last row agree within 0·5 % with f/R_0 and z_1/R_0 values obtained by trajectory computation for the measured field distribution given by curve 2 in fig. 5.

The influence of these effects on the lens characteristics is the more pronounced the shorter the relative gap length S/D, as shown in table 1. For the 'medium' case $S/D=1$ the correction of the dimensions of an 'idealized' lens due to field broadening in working out the imaging properties is negligible for $\mu/\mu_0\geqslant 200$; for $\mu/\mu_0=100$ this correction leads to an increase of the focal length by about 1 %, whereas this increase is about 3 % for $\mu/\mu_0=50$. The main effect here is the loss of m.m.f. in the iron, which can be calculated from the length of the iron circuit and the flux density in the iron, neglecting the increase of flux density at the pole piece tips (in well designed 'real' lenses the flux density in the iron drops with distance from the air gap, according to Mulvey, and the loss of m.m.f. will be unimportant as further discussed below).

As a quantitative illustration of the application of the foregoing results to a practical case, the focal lengths of two electron microscope objectives, one with $S/D=0{\cdot}6$ and one with $S/D=1$, will be worked out from figs. 2 and 15 of I, under conditions of incipient iron saturation of the pole piece tips. The following dimensions are assumed: lens bore $D=3{\cdot}0$ mm, gap length $S=1{\cdot}8$ mm and 3·0 mm respectively, length of each pole piece $L=40$ mm. The iron is assumed to be 'commercial' soft iron (full lines, fig. 3), and the accelerating voltage $V=75$ kv ($V_r=81$ kv); the cross section of the iron circuit is assumed to be indefinitely large, as in fig. 2. As before, a value of μ/μ_0 is chosen, the corresponding value of H is read from fig. 3, and $H_p=(\mu/\mu_0)H$ is found. From H and H_p and the dimensions of the lens the number of ampere turns $(NI)_G$ across the air gap and the total number of ampere turns $(NI)_T$ for the lens are obtained. From $(NI)_G$, V_r and fig. 2 of I the excitation parameter $k^2=\beta\,(NI)_G^2/V_r$ is determined, and with this f_0/R_0, hence f is found from fig. 15 of I. For such high values of μ/μ_0 that field broadening has begun, the equivalent change of S and D has to be taken into account in reading off β from fig. 2 of I and in the subsequent evaluation of f. The result is plotted in fig. 8. The two curves marked 1a and 1b refer to the lens $S/D=0{\cdot}6$, the two curves 2a and 2b to $S/D=1$. Curves 1a and 2a represent the focal lengths in millimetres of the two lenses, assuming $\mu/\mu_0=\infty$, whereas for the curves 1b and 2b account had been taken of the effects of the drop of m.m.f. in the iron of the pole pieces at high excitations and of iron saturation at the pole tips according to table 2. The major part of the difference in the focal lengths in this example is due to the drop in m.m.f. The arrows mark the excitation values where field broadening begins, μ/μ_0 having fallen to approximately 200 for $S/D=0{\cdot}6$ and to approximately 120 for $S/D=1$, with consequent increase in f by 1%. At an excitation of 6000 ampere turns field broadening has increased the focal length by 4% for lens $S/D=1$, but by about 18% for lens $S/D=0{\cdot}6$. One sees from this example that at very high flux densities the focal length curves are considerably modified, the more so the shorter the gap length. This is in agreement with the experimental results of Ruska (1944).

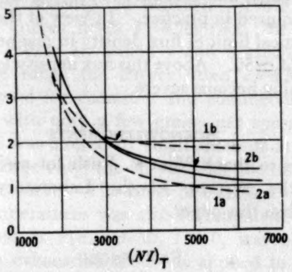


Fig. 8. Computed focal lengths of two lenses as function of lens excitation. Curve (1a): 'ideal' lens $S/D=0{\cdot}6$ ($\mu/\mu_0=\infty$). Curve (1b): 'real' lens $S/D=0{\cdot}6$ (μ/μ_0 as in fig. 3). Curve (2a): 'ideal' lens $S/D=1$ ($\mu/\mu_0=\infty$). Curve (2b): 'real' lens $S/D=1$ (μ/μ_0 as in fig. 3). Arrows mark the excitation values where pole tip saturation effects (field broadening) become noticeable. $V=75$ kv, $D=3{\cdot}0$ mm, $2L=80$ mm.

In the greater part of the foregoing dicussion (except in § 4) the changes in the lens properties due to finite permeability had been considered under the assumption of an indefinitely large iron cross section. This implied a uniform flux density in the iron at a long distance from the air gap. Mulvey's (1953) measurements have shown that this assumption is not justified. In a badly designed iron circuit the flux density increases with distance from the gap (see for example Mulvey 1953, fig. 6, curve for $D_2/D_1=2$), and when saturation effects have to be taken into account it is then mainly the roots of the pole pieces and not their tips which saturate and thus limit lens performance. In a well designed iron circuit (Mulvey 1953, fig. 6, curve for $D_2/D_1=3$) the flux density drops with distance from the air gap before rising again. The average value of flux density in the iron is then so low that the relative reduction in the m.m.f. developed across the air gap,

$$\frac{\delta\Psi}{\Psi}=\frac{2}{S}\int_{S/2}^{S/2+L}\frac{dl}{\mu/\mu_0}, \qquad \ldots\ldots(4)$$

is small owing to the rapid rise of μ/μ_0 with decrease in B, as seen from the right-hand side of fig. 3. In this case field broadening at the pole tips at high flux densities will be the main limit to lens performance.

§ 6. IMAGING PROPERTIES OF LENSES UNDER INCIPIENT IRON SATURATION

In § 4 it was shown that iron saturation at the pole tips can be described by (i) a reduction in the effective m.m.f. across the air gap ('loss of ampere turns') with the simultaneous appearance of a weak parallel field in the lens bore, and by (ii) an apparent increase in the physical dimensions S and D of the lens. Which of these effects is the more important will depend on the design of the iron circuit. In an 'idealized' lens, with indefinitely large iron cross section, the first of these two effects is by far the stronger at the onset of saturation; the effect of the parallel field in the lens bore is negligible as the refracting power is proportional to $\int H^2 dz$. This is illustrated by the example $S/D=1$, $\mu/\mu_0=50$, with such values of S, D_0, NI and V_r that the lens excitation parameter $k^2=0{\cdot}022\,H_0^{\,2}R^2/V_r=1$ for $\mu/\mu_0\to\infty$. From the data given in I and in § 4 of this paper one obtains the data of table 2.

Table 2

No.	μ	f/R_0	z_1/R_0	Remarks
1	∞	1·09	0·68	—
2	50	1·32	0·99	Corrected for drop of m.m.f. across gap
3	50	1·36	1·02	As 2, and corrected for field broadening
4	50	1·35	1·01	As 3, and corrected for parallel field in bore

The data given in the last row agree within 0·5% with f/R_0 and z_1/R_0 values obtained by trajectory computation for the measured field distribution given by curve 2 in fig. 5.

The influence of these effects on the lens characteristics is the more pronounced the shorter the relative gap length S/D, as shown in table 1. For the 'medium' case $S/D=1$ the correction of the dimensions of an 'idealized' lens due to field broadening in working out the imaging properties is negligible for $\mu/\mu_0\geqslant 200$; for $\mu/\mu_0=100$ this correction leads to an increase of the focal length by about 1%, whereas this increase is about 3% for $\mu/\mu_0=50$. The main effect here is the loss of m.m.f. in the iron, which can be calculated from the length of the iron circuit and the flux density in the iron, neglecting the increase of flux density at the pole piece tips (in well designed 'real' lenses the flux density in the iron drops with distance from the air gap, according to Mulvey, and the loss of m.m.f. will be unimportant as further discussed below).

As a quantitative illustration of the application of the foregoing results to a practical case, the focal lengths of two electron microscope objectives, one with $S/D=0{\cdot}6$ and one with $S/D=1$, will be worked out from figs. 2 and 15 of I, under conditions of incipient iron saturation of the pole piece tips. The following dimensions are assumed: lens bore $D=3{\cdot}0$ mm, gap length $S=1{\cdot}8$ mm and 3·0 mm respectively, length of each pole piece $L=40$ mm. The iron is assumed to be 'commercial' soft iron (full lines, fig. 3), and the accelerating voltage $V=75$ kv ($V_r=81$ kv); the cross section of the iron circuit is assumed to be indefinitely large, as in fig. 2. As before, a value of μ/μ_0 is chosen, the corresponding value of H is read from fig. 3, and $H_p=(\mu/\mu_0)H$ is found. From H and H_p and the dimensions of the lens the number of ampere turns $(NI)_G$ across the air gap and the total number of ampere turns $(NI)_T$ for the lens are obtained. From $(NI)_G$, V_r and fig. 2 of I the excitation parameter $k^2=\beta\,(NI)_G^{\,2}/V_r$ is determined, and with this f_0/R_0, hence f is found from fig. 15 of I. For such high values of μ/μ_0 that field broadening has begun, the equivalent change of S and D has to be taken into account in reading off β from fig. 2 of I and in the subsequent evaluation of f. The result is plotted in fig. 8. The two curves marked 1a and 1b refer to the lens $S/D=0{\cdot}6$, the two curves 2a and 2b to $S/D=1$. Curves 1a and 2a represent the focal lengths in millimetres of the two lenses, assuming $\mu/\mu_0=\infty$, whereas for the curves 1b and 2b account had been taken of the effects of the drop of m.m.f. in the iron of the pole pieces at high excitations and of iron saturation at the pole tips according to table 2. The major part of the difference in the focal lengths in this example is due to the drop in m.m.f. The arrows mark the excitation values where field broadening begins, μ/μ_0 having fallen to approximately 200 for $S/D=0{\cdot}6$ and to approximately 120 for $S/D=1$, with consequent increase in f by 1%. At an excitation of 6000 ampere turns field broadening has increased the focal length by 4% for lens $S/D=1$, but by about 18% for lens $S/D=0{\cdot}6$. One sees from this example that at very high flux densities the focal length curves are considerably modified, the more so the shorter the gap length. This is in agreement with the experimental results of Ruska (1944).

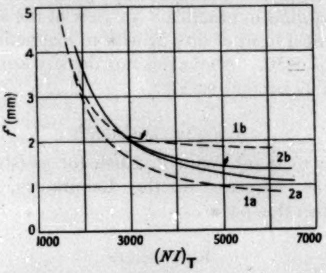


Fig. 8. Computed focal lengths of two lenses as function of lens excitation. Curve (1a) : 'ideal' lens $S/D=0{\cdot}6$ ($\mu/\mu_0=\infty$). Curve (1b) : 'real' lens $S/D=0{\cdot}6$ (μ/μ_0 as in fig. 3). Curve (2a) : 'ideal' lens $S/D=1$ ($\mu/\mu_0=\infty$). Curve (2b) : 'real' lens $S/D=1$ (μ/μ_0 as in fig. 3). Arrows mark the excitation values where pole tip saturation effects (field broadening) become noticeable. $V=75$ kv, $D=3{\cdot}0$ mm, $2L=80$ mm.

In the greater part of the foregoing dicussion (except in § 4) the changes in the lens properties due to finite permeability had been considered under the assumption of an indefinitely large iron cross section. This implied a uniform flux density in the iron at a long distance from the air gap. Mulvey's (1953) measurements have shown that this assumption is not justified. In a badly designed iron circuit the flux density increases with distance from the gap (see for example Mulvey 1953, fig. 6, curve for $D_2/D_1=2$), and when saturation effects have to be taken into account it is then mainly the roots of the pole pieces and not their tips which saturate and thus limit lens performance. In a well designed iron circuit (Mulvey 1953, fig. 6, curve for $D_2/D_1=3$) the flux density drops with distance from the air gap before rising again. The average value of flux density in the iron is then so low that the relative reduction in the m.m.f. developed across the air gap,

$$\frac{\delta\Psi}{\Psi}=\frac{2}{S}\int_{S/2}^{S/2+L}\frac{dl}{\mu/\mu_0}, \qquad \ldots\ldots(4)$$

is small owing to the rapid rise of μ/μ_0 with decrease in B, as seen from the right-hand side of fig. 3. In this case field broadening at the pole tips at high flux densities will be the main limit to lens performance.

The way to determine lens characteristics under conditions of high flux density in practical cases is therefore to consult Mulvey's paper on the expected flux distribution and work out from this (*a*) the loss of ampere turns $\delta\Psi/\Psi$ in the iron, (*b*) the *average* flux density in the pole tips (without the flux increase at the edge of the bore), which is numerically equal to H_p used in I, and hence the value of μ/μ_0 to be used. From this, and the value of S/D, the apparent change in S and in D can be determined from the data of this paper. Then the lens properties can be worked out from I, as shown in the preceding numerical examples.

In conclusion, one can say that for average flux densities in the pole tips corresponding to $\mu/\mu_0 \geqslant 100$ the lens data of 'ideal lenses' as published in I can be used if the iron circuit is properly designed. This value of μ/μ_0 corresponds to 18 kilogauss for 'commercial' iron and to 22·5 kilogauss for 50/50 cobalt–iron. Saturation effects will become progressively noticeable if the flux density is further increased, but can be taken into account in applying the available lens data as far as will be normally required in practice. In view of the sharp drop of μ/μ_0 with increasing B, the practical limit of flux density in magnetic electron lenses is that flux value for which $\mu/\mu_0 \simeq 50$. Above this flux density loss of ampere turns in the iron and field broadening become severe.

ACKNOWLEDGMENTS

The author wishes to thank Miss P. Rush for assistance with the network measurements and the Director of the Laboratory, Dr. T. E. Allibone, for permission to publish this paper.

REFERENCES

DOSSE, J., 1941, *Z. Phys.*, **117,** 437.
HESSE, M. B., 1950, *Proc. Phys. Soc.* B, **63,** 386.
LENZ, F., 1950 a, *Z. angew. Phys.*, **2,** 448; 1950 b, *Optik*, **7,** 243.
LIEBMANN, G., 1950, *Brit. J. Appl. Phys.*, **1,** 92.
LIEBMANN, G., and GRAD, E. M., 1951, *Proc. Phys. Soc.* B, **64,** 956.
MASSACHUSETTS INSTITUTE OF TECHNOLOGY, 1943, *Magnetic Circuits and Transformers* (New York : John Wiley), p. 25.
VAN MENTS, M., and LE POOLE, J. B., 1947, *Appl. Sci. Res.* B, **1,** 3.
MULVEY, T., 1953, *Proc. Phys. Soc.* B, **66,** 441.
RUSKA, E., 1944, *Arch. Elektrotech.*, **38,** 102.

Transient Creep in Pure Metals

By O. H. WYATT
The Cavendish Laboratory, Cambridge*

MS. received 14th July 1952, and in final form 12th January 1953

Abstract. A constant stress testing machine of small inertia and with continuous strain recording has been built. The loading was by air contained in metallic bellows whose wall stiffness was compensated by a 'negative spring'. The initial strain rate was standardized at 1% per second. Polycrystalline copper and aluminium have been investigated between −196°c and 140°c. At low temperatures, the creep curves fitted $\epsilon = \alpha \log t + c_1$; at higher temperatures they fitted $\epsilon = \beta t^{1/3} + c_2$. These are called α and β creep respectively. In an intermediate temperature range the curves fitted $\epsilon = \alpha \log t + \beta t^{1/3} + c_3$. The results have been confirmed for cadmium and commercially pure aluminium, and also for aluminium with only a few grains per specimen. In a series of increment tests the stress was altered by small amounts at various times during a creep curve; in the α range only, the curves after the increments were similar to segments of the parent curve. The effect of testing a specimen at increasing stresses and different temperatures was also investigated; in the α range only, an equation of state, that is $F(\sigma, \epsilon, d\epsilon/dt, T) = 0$, was obeyed. A theory of α creep is developed: an exhaustion theory is applied to a metal whose strain hardening is defined by a reference curve. The rise of activation stress during a creep curve and the slope of the reference curve are deduced from the increment and creep tests; and the slope agrees numerically with the stress–strain curve at −196°c. From an equation of state, values of the mean activation energy at a strain rate of 10^{-4} cm/cm/sec, and the frequency of thermal fluctuations are obtained.

---

§ 1. Introduction

When an annealed single metal is loaded, the deformation continues after the load is applied; at first, the rate of deformation decreases, and finally, at high temperatures, it reaches a constant value. This paper reports an experimental and theoretical investigation of the decelerating or transient creep; a brief report has appeared elsewhere (Wyatt 1951).

It has been generally accepted in this country (Andrade 1910, Orowan 1947) that the transient creep curve is given by

$$\epsilon = \beta t^{1/3} + c_1. \qquad \ldots\ldots(1)$$

However, others (Phillips 1905, Chevenard 1934, Laurent and Eudier 1950) have used

$$\epsilon = \alpha \log t + c_2. \qquad \ldots\ldots(2)$$

In this investigation, the creep curves of wire specimens under constant stress have been measured, using a new machine with pneumatic loading and continuous strain recording. The load could be applied rapidly and readings obtained within a second.

---

* Now at Imperial Chemical Industries, Billingham, Co. Durham.

The main investigation has been with polycrystalline high conductivity copper and polycrystalline spectroscopic purity aluminium. A few tests have been made with three grades of commercial purity aluminium, spectroscopic purity cadmium, and aluminium with only a few grains per specimen.

It was found that, for each metal, the creep curve at low temperatures was given by eqn. (2) (α creep), and, at higher temperatures, by eqn. (1) (β creep). The transition temperature range was lowered by increased stress: thus, for copper the transition with 10 kg mm^{-2} occurred at about 170°c, and with 26 kg mm^{-2} at about 30°c.

Theories of transient creep have been given by Orowan (1947) and Smith (1948) who derived eqns. (1) and (2) respectively. Orowan reported some experiments he carried out with Los which demonstrated that an equation of state, that is $F(\sigma, \epsilon, d\epsilon/dt, T)=0$ did not exist. He emphasized that any theory that led to an equation of state must be wrong.

In this investigation, a series of increment and decrement experiments were made in which the applied stress was increased and decreased a small amount at various times during the creep. Also, the temperature was changed between tests on one specimen at increasing stresses. These experiments demonstrated that an equation of state existed in the α range, but not in the β range.

A theory has been developed by applying the exhaustion theory of Smith to a strain hardening material with a reference stress–strain curve as suggested by Orowan. An equation of state is implicit and has been derived. The theory leads to a numerical relation between the shape of the creep curve and the slope of the stress–strain curve at low temperatures which is satisfied by the observations. Values for the mean activation energy for a given strain or creep rate and the value of the frequency of thermal fluctuations are obtained.

Part I. EXPERIMENTS

§ 2. Apparatus

The principle of the machine is illustrated in fig. 1. The specimen occupies the full length between a piston and one end of the cylinder, in which the piston moves without friction. The air pressure on the specimen side is set to give the required stress; the other side is evacuated. During creep, if the usual assumptions are made that the volume of the specimen remains constant and the reduction of area is occurring uniformly along the specimen, a constant true stress on the specimen is maintained by the isothermal expansion of the air according to the perfect gas law.

Pneumatic loading gave a machine with small inertia. The mass of the moving parts in the actual design was 0·6 kg and the maximum load was 22·5 kg. Hence, the machine was suitable for investigating the initial part of the creep curve, and for the increment experiments, when the strain rates altered rapidly and inertia effects might have masked the true creep curve.

The design is shown schematically in fig. 2, and a photograph of the apparatus in fig. 3. The piston and rigid cylinder were replaced by two elastic bellows, whose wall stiffness was counterbalanced by a negative spring, and the specimen was placed externally.

The load was obtained by setting the air pressure in the lower bellows to between zero and one atmosphere absolute (at higher pressures the bellows became unstable and bowed as a strut); the upper bellows were permanently evacuated. The force of the air pressure on the diaphragm dividing the bellows was transmitted by the loading frame to one end of the specimen, the other end of which was fixed to the main frame. The specimen was gripped in axially loading shackles of the rolling ball type.

The negative spring is shown schematically in fig. 4. It consisted of six equal links, freely joined at their ends, with a helical spring placed across the diagonal AB. The unloaded length of this spring must be equal to links X and Y (these two links need not be the same length as the others). The forces *P* required to maintain equilibrium can be shown to be proportional to the separation of links X and Y, where the constant of proportionality equals the spring characteristic. The negative spring was laterally unstable and the loading frame was kinematically restrained by ball-bearing guides to move along the axis of the specimen.

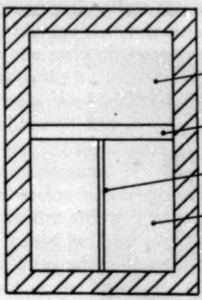


Fig. 1. Principle of constant stress creep testing machine. S = specimen, P = piston, a = air, v = vacuum.

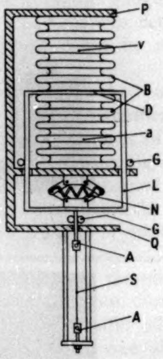


Fig. 2. Scheme of apparatus. A = axial loading shackle, B = bellows, D = diaphragm, G = guides, L = loading frame, N = negative spring, S = specimen, a = air, v = vacuum.

Note : Q as mentioned in the text refers to the centre plate, not the bottom one.

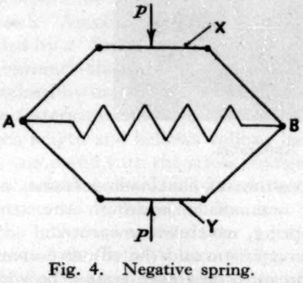


Fig. 4. Negative spring.

To neutralize the stiffness of the bellows the negative spring must be set up so that it is closed (X on Y) when the two bellows are equally compressed, i.e. when the loading frame is not loaded and both bellows are open to the atmosphere.

The range of movement of the central diaphragm was limited because bellows only behave as helical springs when used in compression; thus, for bellows with a free length l_0 and an elastic range d, the distance between plates P and Q in fig. 2 was arranged to be $2l_0-d$. The movement was further limited, to less than d, by axle interference whereby the negative spring operated in one direction only, but an increase, to a limit of d, was achieved by displacing the unloaded position of the loading frame with a light spring attached between the frame and the ceiling. The negative spring was then set up so that theoretically it would close at the new unloaded position of the loading frame, and the characteristic of the diagonal spring was made equal to the combined characteristic of the bellows and light spring, which was measured.

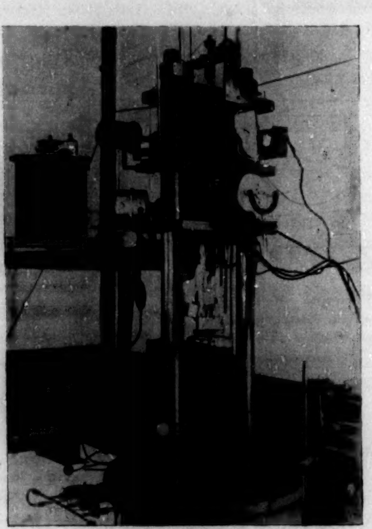


Fig. 3. Apparatus.

It was necessary that the spring characteristic and the effective area of the bellows should be independent of the pressure difference across the walls; the spring characteristics of several makes of bellows were therefore measured, first with both bellows open to the atmosphere and secondly with both evacuated. It was found that bellows with U-shape corrugations were superior to those of V-shape. The thinnest wall to withstand a one-atmosphere pressure difference was chosen to reduce the friction in the negative spring. The bellows used, which were made by the Radio Corporation of America, had an external diameter of 2 in., an internal diameter of $1\frac{3}{4}$ in. and a wall thickness of 0·004 in. The spring characteristic of a single bellows was 1·1 kg/cm.

The length of the specimen was equal to the combined volume of the lower bellows, pipe lines and pressure gauge, divided by the effective area of the diaphragm. This length was found from measurements of the force–position curves of the loading frame with the apparatus in the operating condition, that is, with the upper bellows evacuated and air in the lower bellows. A graph of the inverse of the load plotted against the position of the loading frame was linear and the correct length was deduced from the intercept value on the position axis. The lengths at three different pressures in the pressure range were 9·1, 9·9 and 10·4 cm. This variation did not noticeably affect the linearity of the curve obtained at one pressure because the pressure change was small. Nor did it seriously affect the stress constancy during the flow in a creep test. The stress change $\Delta\sigma$ on a stress σ during strain ϵ was given by $\Delta\sigma/\sigma=(1-r)\epsilon$ where r is the ratio of the correct volume (specimen length times effective area of bellows) to the actual volume. The specimen gauge length was 10·0 cm.

The stress was deduced from readings of the pressure gauge which was directly calibrated and in which the deflection of a bellows was detected by an optical lever. The maximum load was 22·5 kg and could be measured to ± 0·1% of maximum.

The extension was recorded by projecting a graticule attached to the loading frame on to the slit of a rotating drum camera. With a magnification × 200, the strain could be read to 2×10^{-5}. The maximum movement of the loading frame was ½ cm which corresponded to a 5% strain. By turning a nut at the bottom end of the specimen, the loading frame could be returned to the start of its range and further strain accommodated; at the same time the volume was increased 5% to correspond with the new specimen length. Typical recordings are reproduced in fig. 5.

The specimen was surrounded by a Dewar flask. Low temperatures were obtained with liquid nitrogen or methyl alcohol and solid carbon dioxide. Above room temperature the flask was filled with medicinal paraffin oil, and above 140°c with 'Arocolor 1248', heated by two electric coils, one of which was controlled by a 'Sunvic' bimetallic thermostat. The temperature was measured with a chromel–alumel couple with control to ±⅛°c; uniformity of temperature was obtained by use of an electric stirrer.

The systematic error in stress constancy due to incorrect correlation between specimen length and bellows volume had a maximum value $\Delta\sigma/\sigma=0{\cdot}06\ \epsilon$ which may be compared with the stress change in a constant load test $\Delta\sigma/\sigma=\epsilon$. There was also an initial error due to isentropic expansion of the gas, but the expansion rapidly became isothermal. Random fluctuations of the temperature of the air in the bellows were less than ±¼°c, i.e. $\Delta\sigma/\sigma<\pm0{\cdot}08\%$. The friction and hysteresis of the bellows was ± 15 g, which, for an average specimen area of 1·0 mm^2 and a minimum test stress of 6 kg mm^{-2}, gives $\Delta\sigma/\sigma=\pm0{\cdot}25\%$. At higher loads, the percentage error was proportionately less.

§ 3. PREPARATION OF METALS FOR TESTING

The metals tested were as follows:

(*a*) ***High conductivity copper***. The spectrographic analysis of the sample was As 0·0001%, Bi 0·0003%, Fe 0·001%, Pb 0·0003%, Ni 0·0001%, Ag 0·0001%, Sn 0·0001%. The material was received as hard drawn wire in a coil. It was straightened and brought to size by pulling, which extended it about 5%. It was cut to specimen lengths and batches of fifty were annealed *in vacuo* at 420°c for two hours. The grain size was 0·02 mm. The wire diameters were 1·20 mm and, for higher stresses, 0·87 mm; results showed no size effect.

(*b*) *Fine grain spectroscopic purity aluminium.* This sample, which was obtained from Johnson, Matthey and Co. had less than 0·002% impurities. The material was received as hard drawn wire and was prepared as for copper except that annealing was at 400°c *in vacuo* for one hour. The grain size was found to be 0·15 mm. The wire diameter was 2·02 mm.

(*c*) *Large grain spectroscopic purity aluminium.* The material was from the same rod as (*b*). The large grains were produced by the critical strain technique. The specimen cross section was 1·5 mm square; the grains occupied the full cross section and were about 1 cm long.

(*d*) *Commercial purity aluminium in three grades.* The analyses of these samples, which were given by the Aluminium Wire and Cable Co., are shown in table 1. The material was supplied annealed as straight wire of 2 mm diameter.

Table 1. Analyses of three grades of Commercial Purity Aluminium

Grade	Si	Fe	Cu	Mn	Al
99·8%	0·09	0·08	0·002	0·002	Balance
99·5%	0·09	0·16	0·007	0·005	Balance
99·0%	0·24	0·50	Trace	0·02	Balance

(*e*) *Spectroscopic purity cadmium.* This sample was obtained from Johnson, Matthey and Co. Spectrographic examination revealed faintly visible lines of lead, calcium, copper and silver, estimated at less than 0·001% impurity. The material was received hard drawn, straightened by pulling and annealed at 45°c in air for 1½ hours.

§ 4. The Start of the Creep Test

It should be understood that the constant stress creep curve actually includes the period when the stress is being increased from zero up to the test stress. The strain during loading has been called the sudden strain, but this is misleading as the loading cannot be instantaneous. It has arisen by fitting eqn. (1) to later parts of the curve and then extrapolating to zero time.

Preliminary experiments were made to investigate the effect of the loading rate on the creep curve. While the loading frame was held by a stop, the air pressure in the lower bellows was adjusted to give the desired test stress; the stop was then removed at different speeds and recordings were taken. Typical creep curves for copper at 18 kg mm^{-2} and 27°c are given in fig. 5.

The variations between curves at the same stress could be due to the non-existence of an equation of state (Orowan 1947) or to an overshoot effect (Los, unpublished). In the latter, the kinetic energy of the loading frame distorts the creep curve by momentarily loading the specimen above the test stress provided by the air pressure. This effect could not be calculated because the creep theories were uncertain, and an unsuccessful attempt was made to remove the kinetic energy by making the loading frame hit an equal mass. Another method of reducing the overshoot was to remove the stop more slowly, but in the limit the test became a tensile test at constant strain rate. A standard was set at an initial strain rate of 1% per second, at which the overshoot was only

3·5 g mm^{-2}. It is interesting to note the influence of machine inertia: if the stop is removed very rapidly compared with the follow-up rate of the loading frame, the overshoot is independent of the mass of the frame, but at the slower rates used in these experiments the loading frame remains in contact until the stress is fully applied, and the overshoot decreases with reduced mass. Thus, for a given overshoot, the load can be applied more rapidly with this than with a conventional machine.

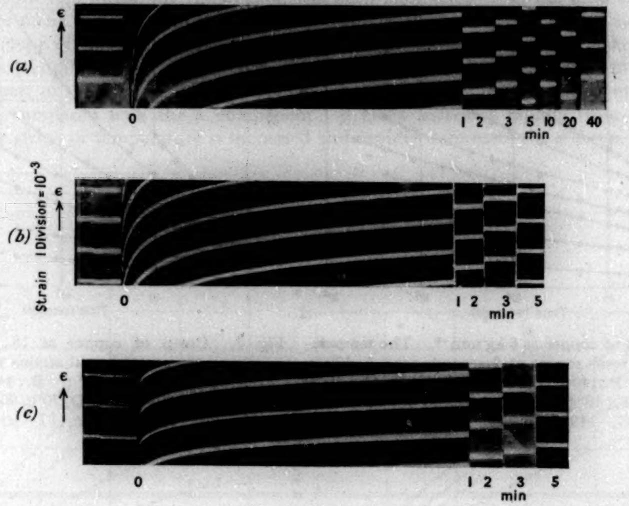


Fig. 5. Recordings of creep curves with copper at 27°c and 18 kg mm^{-2}. The initial strain rates in cm cm^{-1} sec^{-1} and overshoot stresses in g mm^{-2} are: (*a*) 0·01, 3·5; (*b*) 1·0, 350; (*c*) 1·0, 500 (produced by doubling the mass of the loading frame).

§ 5. RESULTS OF CREEP EXPERIMENTS

In these experiments a new specimen was used for each test, and the load was applied in the standard manner described above.

(*a*) *Copper*

Copper was tested with stresses from 6 to 18 kg mm^{-2} over the temperature range -196°c to 170°c, and with stresses up to 38 kg mm^{-2} at -196°c. The standard duration of a test was forty minutes, and the results were confirmed with tests lasting four days. Typical results are given in figs. 6, 7 and 8, in which the total strain is plotted against time on a logarithmic scale. The exact position of zero time is unimportant after a second or two. At low temperatures the graphs are approximately linear and eqn. (2) holds approximately; the actual strain at longer times is less than the theoretical. At the higher temperatures the graphs become concave upwards, especially at longer times. The transition

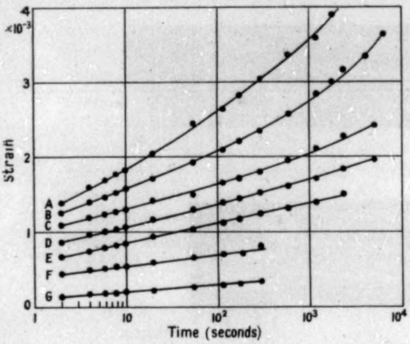


Fig. 6. Creep of copper at 6 kg mm^{-2}. The temperatures and total strains at 2 seconds are: A, 170°c, 15×10^{-3}; B, 140°c, 14×10^{-3}; C, 100°c, 11×10^{-3}; D, 70°c, 11×10^{-3}; E, 27°c, 10×10^{-3}; F, −72°c, 8×10^{-3}; G, −196°c, 6×10^{-3}.

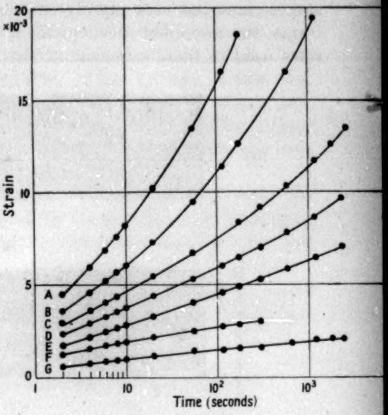


Fig. 7. Creep of copper at 18 kg mm^{-2}. T
temperatures and total strains at 2 seconds a
A, 170°c, 132×10^{-3}; B, 140°c, 116×1
C, 100°c, 99×10^{-3}; D, 70°c, 96×10^{-3}; E, 2
86×10^{-3}; F, −72°c, 72×10^{-3}; G, −19
60×10^{-3}.

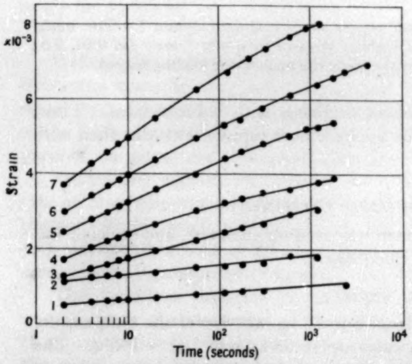


Fig. 8. Creep of copper at −196°c. The stresses in kg mm^{-2}, and total strains at 2 seconds are: (1), 10, 22×10^{-3}; (2) 14, 41×10^{-3}; (3) 22, 89×10^{-3}; (4) 26, 110×10^{-3}; (5) 30, 142×10^{-3}; (6) 34, 172×10^{-3}; (7) 38, 208×10^{-3}.

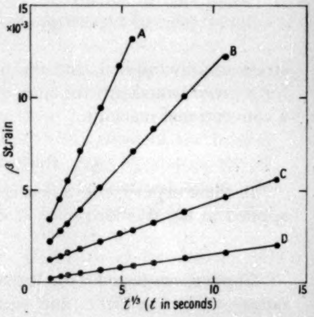
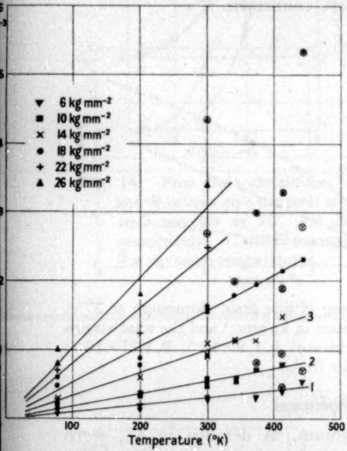


Fig. 9. β components of the creep curves in fig. 7.

temperature range is lower for higher stresses. It was found that the curves fitted an expression of the form

$$\epsilon = \alpha \log t + \beta t^{1/3} + c_3. \qquad \ldots\ldots(3)$$

This is shown in fig. 9 for the curves at 18 kg mm^{-2}; the α component, given by a straight line in the log (time) plot, has been subtracted from the total strain curves of fig. 7 and the difference, the β component, is plotted against (time)$^{1/3}$. The slope of the line representing the α component was adjusted to make the strain–(time)$^{1/3}$ graph linear. The lines were then almost tangential to the first part of the total strain–log (time) graphs, especially if the β component was small compared with the α component. β creep increases rapidly from zero as the stress or temperature is increased and rapidly dominates the α creep.



g. 10. Variation of α with temperature for copper at the different stresses: The encircled points give the mean slopes of the total strain–log(time) curves over the first 100 seconds, i.e. before subtracting the β component.

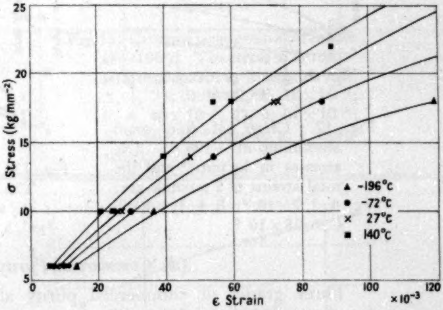


Fig. 11. Stress–strain curves for copper at a strain rate of 10^{-4} cm/cm/second.

The variation of α with temperature and stress is given in fig. 10. When only α creep is present, α is obtained directly from the slopes of the total strain–log (time) graphs. When β creep is also present, it is necessary to subtract the β component from these curves.

In fig. 11, the stress–strain curves at constant strain rate are plotted at various temperatures. These have been taken from the creep curves. That this is permissible will be shown below in the increment experiments.

(*b*) *Fine Grain Aluminium*

Fine grain aluminium was tested at -196°c with stresses from 2 to 6 kg mm^{-2}, and at 27°c and 70°c with lower stresses. The results at -196°c are given in fig. 12, which shows that only α creep was occurring. Results at 27°c are given in fig. 13. Up to 3 kg mm^{-2} the creep curve fits eqn. (1). Some non-linearity occurs initially which may be due to α creep, but β creep is dominant and remains so down to the lowest stress which gave recordable creep. At 5 kg mm^{-2}, which was the highest stress which the machine could accommodate because of the large strains occurring, this plot becomes concave downwards, showing deviations from eqn. (1) which are of the opposite sign to those which would be caused by steady rate creep. Variations of β and initial strain with stress and temperature are given in fig. 14, which shows that these vary rapidly.

(*c*) *Large Grain Aluminium*

Large grain aluminium was tested at −196°c and 27°c (figs. 15 and 16). With this material, β flow was obtained at −196°c by raising the stress sufficiently. Otherwise, the results for fine grain metals were confirmed.

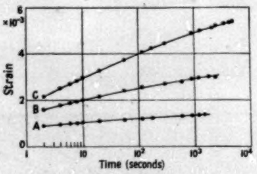


Fig. 12. Creep of fine grain aluminium at −196°c. The stresses in kg mm^{-2} and the total strains at 2 seconds are: A, 2, 7×10^{-3}; B, 4, 21×10^{-3}; C, 6, 38×10^{-3}.

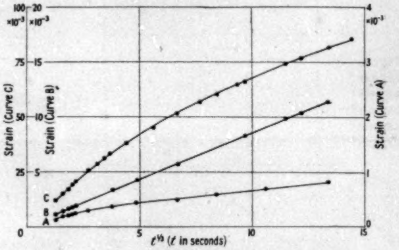


Fig. 13. Creep of fine grain aluminium at 27°c. The stresses in kg mm^{-2} and the total strains at 2 seconds are: A, 2, 9×10^{-3}; B, 3, 27×10^{-3}; C, 5, 112×10^{-3}.

(*d*) *Commercial Purity Aluminium*

Three grades of commercial purity aluminium, as detailed in § 3, were tested at −196°c and 27°c. The creep curves were found to fit eqns. (2) and (1) respectively. The results are tabulated in table 2 for a stress of 6 kg mm^{-2}. It will be seen that impurities have a very much larger effect in the β range.

Table 2. Tests of various grades of Aluminium at 6 kg mm^{-2}

Purity (%)	at −196°c		at 27°c	
	Total strain at 2 sec.	α	Total strain at 2 sec.	β
99·99	0·038	1×10^{-3}	0·230*	10×10^{-3}*
99·8	0·017	$0{\cdot}5\times10^{-3}$	0·033	1×10^{-3}
99·5	0·010	$0{\cdot}3\times10^{-3}$	0·021	$0{\cdot}6\times10^{-3}$
99·0	0·010	$0{\cdot}3\times10^{-3}$	0·023	$0{\cdot}6\times10^{-3}$

* at 5·75 kg mm^{-2}.

(*e*) *Cadmium*

Cadmium was tested as representative of the hexagonal metals. The results are given in fig. 17. At −196°c the curve fits eqn. (2) and the results for face-centred cubic crystals are confirmed. At 27°c and 2 kg mm^{-2} the curve did not fit eqn. (1).

All the materials were tested for recovery creep on unloading, but none was detected. On reloading, copper in the α range continued approximately the creep curve from the point at which it had been terminated by unloading. On reloading aluminium in the β range, the initial total strain and creep rate were

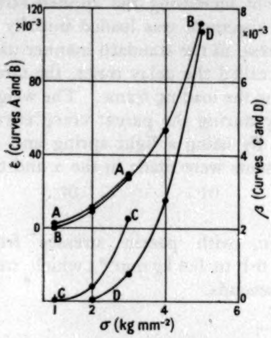


Fig. 14. Fine grain aluminium. A and B curves give the total strain at 2 seconds at 70° and 27°c respectively. C and D curves are β at the same temperatures.

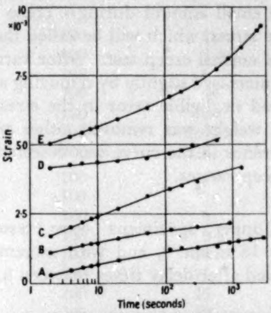


Fig. 15. Creep of large grain aluminium at −196°c. The stress in kg mm⁻² and total strain at 2 seconds are: A, 1, 7×10^{-3}; B, 2, 18×10^{-3}; C, 4, 40×10^{-3}; D, 6, 61×10^{-3}; E, 8, 96×10^{-3} (drawn at ¼ strain scale).

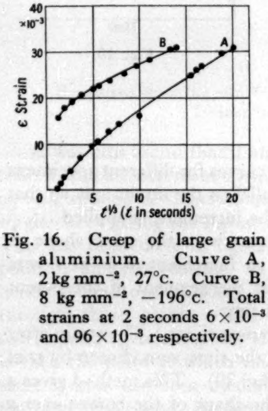


Fig. 16. Creep of large grain aluminium. Curve A, 2 kg mm⁻², 27°c. Curve B, 8 kg mm⁻², −196°c. Total strains at 2 seconds 6×10^{-3} and 96×10^{-3} respectively.

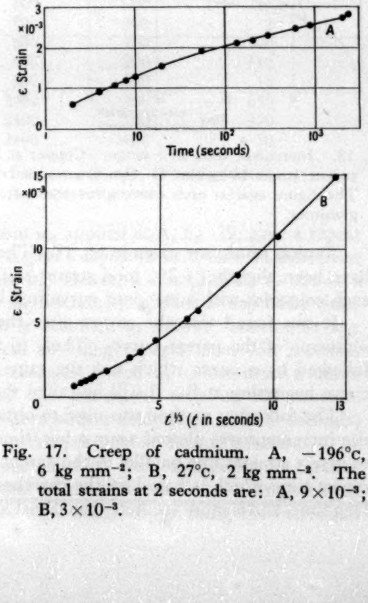


Fig. 17. Creep of cadmium. A, −196°c, 6 kg mm⁻²; B, 27°c, 2 kg mm⁻². The total strains at 2 seconds are: A, 9×10^{-3}; B, 3×10^{-3}.

larger than before unloading, and they increased the longer the material was left unloaded. There was unfortunately not time to investigate this effect fully.

§ 6. INCREMENT EXPERIMENTS

These experiments investigated the effect of increasing the applied stress by a small amount during a creep test. The specimen was loaded initially to some stress, which will be called the parent stress, in the standard manner used for a normal creep test. After various times, called the delay times, the stress was increased slightly by removing a weight from the loading frame. The weight caused negligible error in the stress constancy during the parent creep curve. The weight was removed either suddenly or by using a light spring and no difference in the curve was detected. Experiments were made in the α and the β creep ranges.

α Range

Copper specimens were tested at 27°C, with parent stresses from 6 to 18 kg mm^{-2}, and with increments from 0·1 to 0·4 kg mm^{-2}, which were applied after delay times between 8 and 2400 seconds.

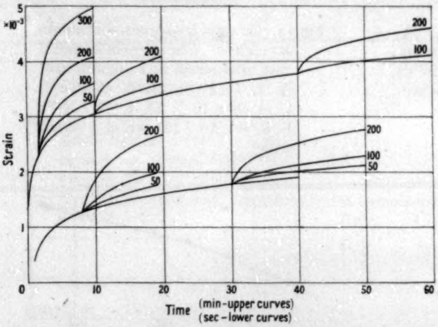


Fig. 18. Increment tests in α range. Copper at 27°C and parent stress 14 kg mm^{-2}. Specimen area 1·12 mm^2. The figure against each curve gives the increment in grammes.

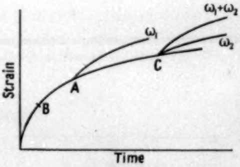


Fig. 19.

Typical results are given in fig. 18. The parent curves for different specimens have been slightly (±2% total strain) shifted parallel to the strain axis so that each coincides with a standard curve just before the increment is applied.

It was found that the curves after the increment were the same shape as segments of the parent curve. Thus, in fig. 19, an increment of ω_1 at A was followed by a curve which was the same shape as the segment of the parent curve beginning at B. B will be called the repeat time.

The following method was used to obtain the repeat times. The strain after the increment was plotted against log (time), with the time zero chosen by trial to give a straight line parallel to the parent curve (fig. 20). This method gives a repeat time which is based on the correlation of the shape of the curves over a long time, and is more accurate than direct fitting of the curves, because sometimes more strain occurred in the first ten seconds after the increment than was required for fitting the curve for, say, the following ten minutes.

The repeat times for various increments and delay times at each parent stress are given in table 3.

Table 3

S	A	D	I	R
6	1·12	15	100	1
		40	100	3·5
		100	100	8
		100	200	3
		600	100	70
		2400	100	200
		2400	200	25
10	1·10	8	100	1
		15	100	3
		40	100	9
		40	200	0·8
		100	100	14
		600	100	75
		600	200	25
		2400	200	70
		2400	300	50
14	1·09	see fig. 18		
18	1·04	8	100	4
		8	200	1·25
		15	100	7
		15	200	2
		40	100	15
		100	100	40
		100	200	17
		100	400	2
		600	100	300
		600	200	110
		600	400	10
		2400	100	650
		2400	200	370
		2400	400	30

S = parent stress (kg mm^{-2}); A = specimen area (mm^2); D = delay time (sec); I = increment (g); R = repeat time (sec).

It was also found that if an increment ω_1 applied at A, fig. 19, gave a repeat time B, and an increment ω_2 applied at C gave a repeat time A, then an increment $\omega_1+\omega_2$ applied at C gave a repeat time B. Hence, it was possible to obtain a single curve giving the repeat time for any increment applied at any delay time. This has been done in fig. 21, which is obtained as follows: starting at any point A at 2400 seconds, an increment of 100 g repeats from 750 seconds, hence B, and one of 200 g from 250 seconds, hence C. Now draw a curve through these points. The zero for the increment tests at 600 seconds (corresponding to the point A for the tests with a delay time of 2400 seconds) is then the point at which this curve intercepts the 600-second abscissa. The same symbol has been used to mark points which were obtained from tests at the same delay time. It can be seen that all the symbols lie on a single curve, which would be so whatever axis scales were used. By plotting the increment against log(time) a linear plot was obtained. The slopes of this line for various parent stresses are given in table 4. The slope increases at higher parent stresses; in other words, the higher the parent stress, the greater the increment required to obtain a given repeat time from a given delay time.

Table 4. Slopes of increment (in kg mm^{-2}) against ln(time) (time in seconds)

Parent stress (kg mm^{-2})	6	10	14	18
Slope $\times 10^3$	3·9	5·5	8·0	10·8
Total strain $\times 10^3$	11	30	55	86

It follows from these results that a creep test at one stress is not affected by a test on the specimen at a lower stress as long as the increment is sufficient to bring the creep rate up to the standard loading strain rate.

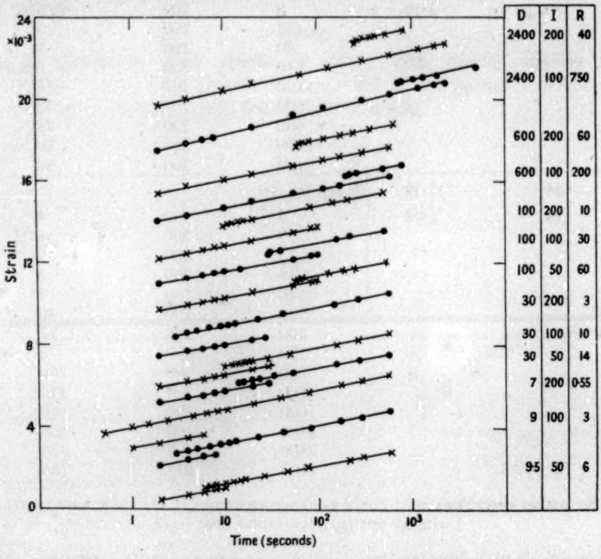


Fig. 20. Obtaining the repeat times for the curves in figure 18. D = delay time (sec); I = increment (g); R = repeat time (sec).

β Range

Fine grain spectroscopic purity aluminium specimens were tested at 27°c, with a parent stress of 3 kg mm^{-2}, and with increments of 0·03 and 0·06 kg mm^{-2} which were applied after delay times between 7 and 2400 seconds. The composite results are given in fig. 22. The same procedure has been adopted of superimposing the parent creep curves just before the increment. The curves are not similar to the parent curves, and no analysis has been achieved. If the increment was large enough to produce an initial creep rate up to the standard loading rate, the curves following the increment fitted equation (1). The value of β and the total strain at 2 seconds depended on the history.

§ 7. Decrement Experiments

The applied stress was decreased by small amounts at various times during a creep test, otherwise the procedure was the same as for the increment tests. Only the α range was investigated, using copper at 27°c, with a parent stress of 10 kg mm^{-2} and decrements from 0·3 to 0·5 kg mm^{-2}.

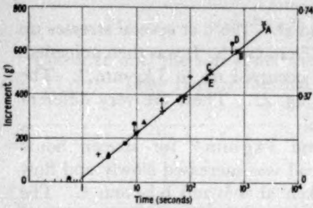


Fig. 21. Increment experiments in the α range. Copper at 27°c and 14 kg mm^{-2}.

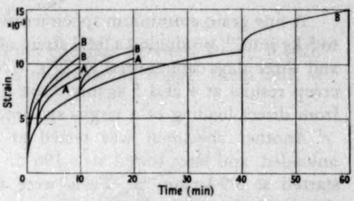


Fig. 22. Increment experiments in the β range. Fine grain aluminium at 27°c and parent stress 3 kg mm^{-2}. Increments of 100 g (A curves) and 200 g (B curves) at delay times of 8, 120, 600 and 2400 sec; specimen area 3·07 mm^2.

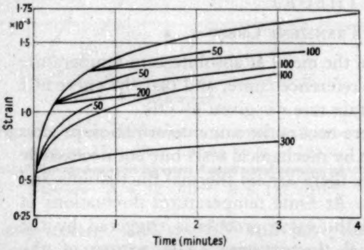


Fig. 23. Decrement experiments in α range. Copper at 27°c and parent stress 10 kg mm^{-2}. Delay times 8, 15 and 40 sec, specimen area 1·10 mm^{-2}. The figure against each curve gives the decrement in grammes.

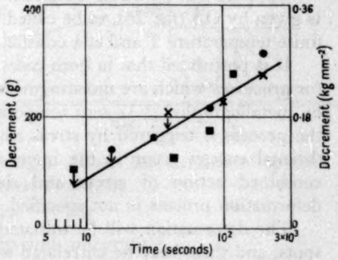


Fig. 24. Decrement experiments in α range.

The composite result of tests on several specimens are given in fig. 23. The curves after the decrement are similar to segments of the parent curves at a later time, and the same analysis as for the increment tests can be made. Figure 24 gives the equivalent of fig. 20, and the slope of this graph is the same as for the increment tests at this stress.

§ 8. Temperature Changed between Tests

A specimen was tested at several stresses at one temperature, unloaded, and its temperature changed. It was then reloaded to a stress at which creep occurred and recordings were taken at this and higher stresses.

α *Range*

A copper specimen was tested at −196°c at several stresses up to 15 kg mm^{-2}, unloaded, and three days later tested at 27°c. The load was increased slowly, flow began at 14 kg mm^{-2}, and tests were taken at 15, 16 and 18 kg mm^{-2}. All the creep curves were identical with those of a directly loaded virgin specimen, that is, the total strain at, say, 2 seconds, and the slope of the strain–log(time) graph were the same. These, and the increment and decrement experiments, show that copper in the α range obeys an equation of state, $F(\sigma, \epsilon, d\epsilon/dt, T) = 0$.

β *Range*

A fine grain aluminium specimen was tested at -196°c at several stresses up to 5 kg mm^{-2}, producing a total strain of 3×10^{-2} cm cm^{-1}. It was then unloaded, and three days later tested at 27°c. No flow occurred up to 3 kg mm^{-2}. The creep results at 4 and 5 kg mm^{-2} are given in fig. 25. These are very different from direct loading of a virgin specimen.

Another specimen was tested at 27°c and 3 kg mm^{-2} for sixteen hours unloaded, and later tested at -196°c. The load was increased slowly and flow started at 3·9 kg mm^{-2}. Tests were then taken at 4·5 and 6 kg mm^{-2}. The slopes of the strain–log(time) graphs were the same as for direct loading, although the total strain at 2 seconds was about double.

Part II. THEORY

§ 9. Theory of α Transient Creep

Suppose that the stress–strain curve of the metal at absolute zero temperature is given by OA (fig. 26), to be called the reference curve, and that the curve at a finite temperature T and at a constant strain rate r is given by OB.

It is postulated that in both cases there occurs the same deformation process (or processes which are indistinguishable by mechanical tests but not necessarily by metallography). At zero temperature there are no thermal fluctuations and the process is triggered by stress alone. At finite temperatures fluctuations of thermal energy occur in the metal and the same process is triggered by the combined action of stress and thermal fluctuations. The nature of the deformation process is not specified.

The deformation will be initiated at many places, which will be called soft spots, and these may be correlated with parts of the reference curve. Suppose soft spot groups x_1, x_2, x_3 and x_4 are associated with sections OD, DF, FG and FA respectively. Then, if stress σ_a is applied at zero temperature, group x_1 will be triggered and the state of the material will be given by D. If the temperature is now increased, spots in groups x_2, x_3 and x_4 will be triggered.

For a given applied stress, every spot has associated with it an activation energy and activation stress which are the minimum additional energy and stress respectively required to trigger a deformation process there. For an applied stress σ_a suppose the distribution of soft spots x_2, x_3 and x_4 against activation stress and energy are as given in fig. 27: the number of spots with activation stress between θ and $\theta + d\theta$ at time t is $N(\theta, t)\, d\theta$; similarly for the activation energy. The broken line gives the initial distribution $N(\theta, 0)$ or $N(E, 0)$ when the metal is in state D. $N(\theta, 0)$ can be deduced from the reference curve, since it equals the product of the slope at $\sigma_a + \theta$ times the contribution to the macroscopic strain of a triggered soft spot. $N(E, 0)$ will then depend on the relation between E and θ. It will be assumed that $N(E, 0)$ is constant with respect to E during a creep test. After a time, some of the spots will have been triggered by the thermal fluctuations and the distribution will be as shown by the continuous line, whose shape will now be calculated.

If x systems are exchanging thermal energy, the number which will have energy between A and $A+dA$ is given by Boltzmann's law $dx = \text{const.}\, e^{-A/\boldsymbol{k}T}\, dA$.

Integrating, we obtain $x = \text{const}\ [e^{-A/\boldsymbol{k}T}]_{\infty}^{0}\, \boldsymbol{k}T$; hence $x/\boldsymbol{k}T = \text{const.}$ At any instant the number of systems with energy A or greater is

$$x_{\geqslant A} = \int_A^{\infty} dx = x e^{-A/\boldsymbol{k}T}.$$

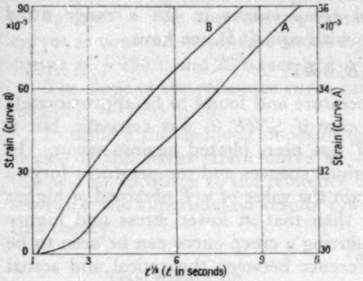


Fig. 25. Creep of fine grain aluminium specimen at 27°c after testing it at −196°c. The stresses in kg mm⁻² and total strains at 2 sec are: A, 4, 32×10^{-3}; B, 5, 111×10^{-3}.

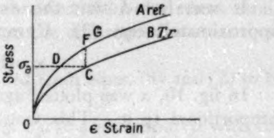


Fig. 26.

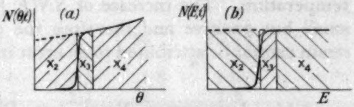


Fig. 27.

Consider now the soft spots with activation energies between E and $E+dE$ (fig. 27 (*b*)). They require thermal energy E or greater to trigger them and, at any instant, $N(E, t)\, e^{-E/\boldsymbol{k}T}$ will have that energy. Then the number of spots which will be triggered in the next time interval dt will be

$$dN(E, t) = -N(E, t)\, e^{-E/\boldsymbol{k}T}\, \nu\, dt$$

where ν is the frequency of thermal vibrations and, integrating,

$$N(E, t) = N(E, 0) \exp\{-e^{-E/\boldsymbol{k}T}\, \nu t\}.$$

Consider the value of E at which $N(E, t)$ is some fraction, say p, of its initial value:

$$p = \exp[\{-\exp(-E_p/\boldsymbol{k}T)\}\nu t]. \qquad \ldots\ldots(4)$$

Hence

$$dE_p/dt = \boldsymbol{k}T/t. \qquad \ldots\ldots(5)$$

This is independent of p, therefore the thermal front remains constant in shape as it advances towards the higher energy spots. We will locate its position by the mean activation energy E_{m} at which $N(E, t) = \frac{1}{2}N(E, 0)$. Similarly for the mean activation stress θ_{m}.

The creep strain rate is given by the product of the spots eliminated per second times the contribution to the macroscopic strain per spot, which will be assumed constant, say S:

$$\frac{d\epsilon}{dt} = SN(E, 0)\,\frac{dE}{dt} = \frac{SN(E, 0)\boldsymbol{k}T}{t} \qquad \ldots\ldots(6\,a)$$

whence

$$\epsilon = SN(E, 0)\boldsymbol{k}T \ln t + \text{const.} \qquad \ldots\ldots(6\,b)$$

It is evident from eqn. (4) that p increases rapidly from almost zero to nearly unity in a small range lying on either side of the mean activation energy. Thus the creep rate is very largely due to the triggering of spots close to E_m and the higher energy spots contribute little. It therefore makes little difference if these are not present and there is only a band of energies present at any time. The higher energy soft spots could then be created by the triggering of these spots. This appears to be physically more realistic than assuming that all the soft spots are present initially. Hence, the theory need not be regarded as a pure exhaustion theory.

§ 10. CREEP CURVES

It was found that the experimental creep curves in the α range fitted approximately eqn. (2). Comparing this with eqn. (6 *b*), we have

$$\alpha = (\ln 10)\, SN(E, 0)\boldsymbol{k}T. \qquad \ldots\ldots(7)$$

In fig. 10, α was plotted against temperature and found to be approximately proportional to it. This would be the case if $SN(E, 0)$ was constant, but it varies with total strain. In fig. 28, α/T has been plotted against strain. In general, the points obtained from the various stresses and temperatures form a continuous curve, although at a given strain the value of α/T obtained at higher stress and lower temperature is greater than that at lower stress and higher temperature. The increase of $SN(E, 0)$ during a creep curve can be seen to be small but positive and therefore the difference between theoretical and actual strain cannot be attributed to an error in the assumption that $SN(E, 0)$ is constant.

§ 11. INCREMENT TESTS AND RELATION BETWEEN CREEP CURVE AND REFERENCE STRESS–STRAIN CURVE

These can be interpreted as evidence of the movement of a thermal front of constant shape. Altering the stress is equivalent to shifting the $N(\theta, t)$ ordinate in fig. 27 (*a*): an increment $\Delta\sigma$ reduces the activation stress of all spots by that amount, and is therefore equivalent to moving the ordinate $\Delta\sigma$ to the right without altering the shape of the curve. This, of course, does not hold for the $N(E, t)$ curve unless E is a linear function of θ. The creep that follows will be the same as occurred during the parent creep curve when the front was the same distance from the ordinate; the total strain has altered so little that $SN(E, 0)$ is constant, compared with the situation with large stress increases. The decrement tests will follow the same pattern in reverse.

Hence the increment tests enable the movement of the front to be measured. Figure 21 shows the rise of mean activation stress during creep. It is given by

$$\theta_m = A \ln t + \text{const.} \qquad \ldots\ldots(8)$$

where A is constant with time and varies slowly with strain, but may be taken as constant during tests at one parent stress. Values of A have been given in table 4.

It will be seen that the greater strains (i.e. after the loading period) at higher stresses are due to the front advancing through a greater stress range, as well as to the decreased slope of the stress–strain curve, to which it has been attributed by Orowan.

Substituting the specific value E_m for E_p in eqn. (5) and integrating we obtain $E_m = \boldsymbol{k}T \ln t + \text{const.}$ Hence by eqn. (8)

$$E_m = \frac{\boldsymbol{k}T}{A}\theta_m + B \qquad \ldots\ldots(9)$$

where A and B are functions of the total strain. Values of A were obtained only at one temperature, but it is probable that A is proportional to T since the relation between E_m and θ_m will not depend on temperature.

Equation (9) should be compared with that due to Becker, which is usually accepted: $E = \text{const. } \theta^2$. Perhaps eqn. (9) is an approximation to this, since it has been deduced from only a small range of θ. Thus, in the increment tests at 18 kg mm^{-2}, θ_m had a value greater than 4 kg mm^{-2}, whilst it altered by only 0·5 kg mm^{-2}. The approximate value of θ_m is obtained from the stress–strain curves at constant strain rate and it is greater than the difference between these curves at −196°c and 27°c (see § 12).

The slope of the reference curve can be deduced from eqns. (8) and (2) to be $(d\theta_m/d\epsilon)_{\sigma_a} = (A/\alpha) \ln 10$.

In table 5 the deduced slopes of the reference curve are compared with the slopes of the stress–strain curve at −196°c and 10^{-4} cm cm^{-1} sec^{-1}, which is the nearest approach to the reference curve obtained experimentally. It will be seen that the experimental results provide a numerical check of the validity of this theory.

Table 5

Total strain $\times 10^3$	10	30	54	86
Reference curve slope (kg mm^{-2}) $\times 10^{-2}$	3·0	2·3	1·8	1·6
Stress–strain curve slope $\times 10^{-2}$	2·7	2·2	2	1·5

These experiments also demonstrate that there is no point in a creep curve at which 'sudden' extension finishes and creep begins. During the loading the thermal front will be just so far ahead of the applied stress that the free creep rate, that is, the initial creep rate if the stress is momentarily held constant, equals the applied strain rate. During loading and subsequently, the deformation process is being triggered by the applied stress and thermal fluctuations. It is therefore misleading to discuss the 'start' of creep.

§ 12. TEMPERATURE CHANGING TESTS

It has already been postulated that the same deformation process occurs at all temperatures (in the α range) and that at different temperatures the thermal fluctuations cause different strains. If a stress is applied, the shape of the thermal front is determined by the temperature, and the value of the mean activation stress θ_m by the temperature and the strain rate. It has already been pointed out that it follows from eqn. (4) that the front between triggered and non-triggered spots occurs in a narrow band of E or θ. It follows that, as a close approximation, the fundamental state of the metal can always be given by a point on the invariable reference curve, irrespective of the temperature at which straining occurred; it is an approximation because the shape of the front depends on the temperature. If the metal is in condition F, fig. 26, through being strained at absolute zero, all the soft spots associated with OF have been triggered and none from FA. On the other hand, if the same total strain be obtained at a finite temperature, some of the OF spots will remain and some of the FA spots will have been triggered; the middle of the front will be centred on F. The distribution curve becomes less square in form as the temperature is raised. But, in the temperature range in which only α creep occurs, it is permissible to consider the fundamental state of the material to be fixed by a point of the reference curve.

This is confirmed by the tests in which the temperature was changed between taking creep curves at increasing applied stresses. It was found that the creep curve was independent of previous straining history as long as additional straining occurred at the start; this is sufficient to eliminate any differences due to the different shape of the front at different temperatures.

§ 13. EQUATION OF STATE

This theory of transient creep leads to an equation of state, $F(\sigma, \epsilon, d\epsilon/dt, T)=0$, which was found to exist in the experiments.

Let the reference curve be given by $\sigma_{\mathrm{r}}=R(\epsilon)$. Following the discussion of the previous section, the applied stress is given by $\sigma_{\mathrm{a}}=\sigma_{\mathrm{r}}-\theta_{\mathrm{m}}$.

Therefore

$$\sigma_{\mathrm{a}}=R(\epsilon)-\theta_{\mathrm{m}} \qquad \ldots\ldots(10)$$

where θ_{m} is the mean activation stress at which half the spots are triggered. θ_{m} is a function of ϵ, $d\epsilon/dt$ and T. From eqn. (4), putting $p=\frac{1}{2}$ and $E=E_{\mathrm{m}}$

$$\ln 2=\nu t \exp\left(-E_{\mathrm{m}}/\boldsymbol{k}T\right).$$

From eqn. (6 *a*)

$$t=SN(E, 0)\boldsymbol{k}T/(d\epsilon/dt).$$

Hence substituting

$$\frac{(d\epsilon/dt)\ln 2}{\nu SN(E, 0)\boldsymbol{k}T}=\exp\left(-E_{\mathrm{m}}/\boldsymbol{k}T\right)$$

i.e.

$$E_{\mathrm{m}}=\boldsymbol{k}T\ln\left\{\frac{\nu SN(E, 0)\boldsymbol{k}T}{(d\epsilon/dt)\ln 2}\right\}.$$

By eqn. (9), writing $A/\boldsymbol{k}T=C_1$, $\theta_{\mathrm{m}}=C_1E_{\mathrm{m}}+C_2$ where C_1 and C_2 are functions of the total strain ϵ.

Hence

$$\theta_{\mathrm{m}}=C_1\boldsymbol{k}T\ln\left\{\frac{\nu SN(E, 0)\boldsymbol{k}T}{(d\epsilon/dt)\ln 2}\right\}+C_2.$$

Substituting in eqn. (10)

$$\sigma_{\mathrm{a}}=R(\epsilon)+C_1\boldsymbol{k}T\ln\{(d\epsilon/dt)\ln 2\}-C_1\boldsymbol{k}T\ln\{\nu SN(E, 0)\boldsymbol{k}T\}+C_2. \qquad \ldots\ldots(11)$$

where C_1, C_2, ν, S, $N(E, 0)$ are functions of the total strain only.

§ 14. VALUES OF ν AND E_{m}

Consider the variation of the applied stress with temperature for a fixed strain and strain rate. Rearranging eqn. (11):

$$T\ln T+\frac{\sigma_{\mathrm{a}}}{C_1\boldsymbol{k}}=\frac{R(\epsilon)+C_2}{C_1\boldsymbol{k}}+\left[\ln\{(d\epsilon/dt)\ln 2\}-\ln\{\nu SN(E, 0)\boldsymbol{k}\}\right]T. \quad \ldots(11\,a)$$

Hence, if $(T\ln T+\sigma_{\mathrm{a}}/C_1\boldsymbol{k})$ is plotted against T, the points should be on a straight line. This is shown in fig. 29 for copper at various strains; $C_1\boldsymbol{k}$ was deduced from the increment tests, since $C_1\boldsymbol{k}=A/T$ and A is defined by eqn. (8). The slope equals $[\ln\{(d\epsilon/dt)\ln 2\}-\ln(\nu SN(E, 0)\boldsymbol{k}]$. Inserting the value of $SN(E, 0)$

which by eqn. (7) may be deduced from the creep curves, the frequency of thermal fluctuation is obtained; values are given in table 6. These are considerably higher than those usually accepted in dislocation theory. The values are dependent exponentially on the terms in eqn. (11 *a*). Hence any approximation in the earlier steps of this theory will produce a large error in ν. One such approximation is the relationship between E_m and θ_m which has been discussed in §11.

Table 6. Frequency of Thermal Fluctuations in Copper

% Strain	1·5	3	6	10
ν	5×10^{18}	$5{\cdot}5\times10^{19}$	6×10^{20}	2×10^{22}

Table 7. Activation Energies at Strain Rate 10^{-4} sec^{-1}

Strain $\times 10^2$	1·5	3	60	100
$E_m/\boldsymbol{k}T$ (ev)	38·7	41	43·3	46·8
E_m at 27° c (ev)	1	1·06	1·12	1·21

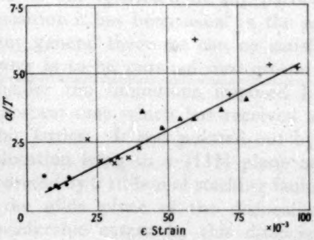


Fig. 28.

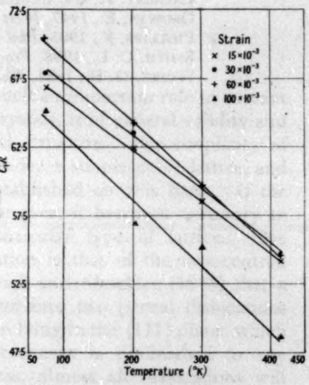


Fig. 29. Copper at strain rate 10^{-4} cm cm^{-1} sec^{-1}.

E_m is obtained from eqns. (4) and (6 *a*) with $p=\frac{1}{2}$, whence

$$\frac{d\epsilon}{dt}=\frac{SN(E,0)\boldsymbol{k}T}{\ln 2}\exp\left(-E_m/\boldsymbol{k}T\right)=\frac{\alpha\nu\exp\left(-E_m/\boldsymbol{k}t\right)}{\ln 10\ln 2}.$$

Values of E_m are given in table 7.

§ 15. CONCLUSIONS

There are two types of transient creep in pure metals. At low temperatures and stresses, α creep predominates and the strain is proportional to log (time); at high temperatures and stresses, β creep predominates and the strain is proportional to (time)$^{1/3}$. In an intermediate range the creep is compound. In the α range only, when the stress and temperature are altered in a specimen, the metal obeys an equation of state $F(\sigma, \epsilon, d\epsilon/dt, T)=0$.

α creep and the experiments in the α range can be explained by a theory in which the deformation occurs by the thermal activation of soft spots whose distribution is represented macroscopically by a single stress–strain curve or reference curve. All creep, speed and temperature effects can then be considered in terms of the shape and position of the front dividing activated from non-activated spots. β creep may be related to some recovery process.

ACKNOWLEDGMENTS

My thanks are due to Professor Sir Lawrence Bragg for facilities at the Cavendish Laboratory, to Dr. E. Orowan for suggesting this investigation and the principle of the machine, together with his ideas on the theory of creep, and finally to Mr. F. R. N. Nabarro for criticizing this paper in draft.

REFERENCES

ANDRADE, E. N. da C., 1910, *Proc. Roy. Soc.* A, **84,** 1.
CHEVENARD, P., 1934, *Rev. Métall.* **31,** 473.
LAURENT, P., and EUDIER, M., 1950, *Rev. Métall.*, **47,** 39.
OROWAN, E., 1947, *West of Scotland Iron and Steel Inst.*, p. 45.
PHILLIPS, F., 1905, *Phil. Mag.*, **9,** 513.
SMITH, C. L., 1948, *Proc. Phys. Soc.*, **61,** 201.
WYATT, O. H., 1951, *Nature, Lond.*, **167,** 866.

Dislocation Nodes in Face-Centred Cubic Lattices

By N. THOMPSON
H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott; *MS. received* 28*th January* 1953

Abstract. Two or more dislocation lines meeting at a point constitute a node. In face-centred cubic lattices a dislocation can split into two partial dislocations. The requirements of crystal geometry impose certain limitations on the arrangement of partial dislocations at a node. The various possibilities are discussed. It is shown that some nodes are fixed relative to the lattice, while others can move by glide processes in certain crystallographic directions. The length of dislocation line joining two nodes can function as a Frank–Read source. The effect of the node structure on the behaviour of the source is considered, and it is shown that such a source may progressively change its character during the process of dislocation production.

---

§ 1. INTRODUCTION

THE concept of a dislocation, which plays such an important role in current theories of the mechanical properties of crystals, is of general validity and is not restricted to any particular crystal structure. For simplicity of exposition it has been usual in the past to consider a simple cubic lattice, and many general theorems can be satisfactorily established on this basis. If the matter is to be pursued into greater detail, however, it becomes necessary to consider the limitations imposed by each particular type of lattice. One important case which has received some attention is that of the face-centred cubic lattice. It was pointed out by Heidenreich and Shockley (1948) that a dislocation lying in a $\{111\}$ plane can dissociate into two partial dislocations separated by a ribbon of stacking fault, the ribbon lying in the $\{111\}$ plane, which is the glide plane of the dislocation. If the energy is diminished to any considerable extent by this dissociation process, almost all dislocations will tend to lie in $\{111\}$ planes, and be so dissociated. This idea provides a natural basis for an explanation of the almost universal occurrence of $\{111\}$ planes as glide planes in face-centred cubic crystals.

There are four essentially different sets of $\{111\}$ planes, lying parallel respectively to the four faces of a regular tetrahedron, and the ribbons of stacking fault, which can be regarded as constituting the extended dislocations, must be thought of as confined to these planes. One obtains a very convenient nomenclature by using such an imaginary tetrahedron as a basis of reference. Corresponding to the fact that each atom has twelve nearest neighbours, there are twelve possible Burgers vectors for the simplest perfect (i.e. undissociated) dislocation. These are all equal in magnitude ($=b$, say) and are representable by the six edges of the same tetrahedron, each considered in two senses. If a perfect dislocation dissociates into two partials, the Burgers vectors of the partials will lie in the glide plane, and will be representable by the two lines which join the ends of the appropriate edge of the tetrahedron to the mid-point of that face which is parallel to the glide plane.

Figure 1 represents the four faces of the tetrahedron opened out so as to lie in one plane. The vertex D was originally below the plane of the paper. The following nomenclature will be used. Vertices of the tetrahedron will be denoted by A, B, C, D, the mid-points of the opposite faces by α, β, γ, δ respectively, and the faces in which these points lie by (a), (b), (c), (d). Burgers vectors will be specified by their two end points, thus: **AB**, **A**$\boldsymbol{\beta}$.

If several dislocation lines meet at a point it is geometrically necessary that the sum of their Burgers vectors be zero (Frank 1951). The vectors must be defined in some consistent manner from the viewpoint of an observer, situated at the point in question, and looking outwards along each dislocation line in turn.* In a face-centred cubic lattice this condition alone clearly permits 2, 3, 4 or—in principle—any number of perfect dislocation lines to meet at a point, forming a 2-fold, 3-fold, etc. node. So long as the dislocations are dissociated into partials, they will be constrained to lie in {111} planes, but no restriction is placed on their directions in those planes. The strain energy in the surrounding crystal can be regarded as conferring on the dislocations a line tension. Thus, in the absence of any crystallographic restrictions on their positions, three dislocation lines meeting at a point in an otherwise perfect crystal subject to no externally applied stress would be straight, would lie in one plane and would make angles of 120° with one another. In a well-annealed real crystal, even when any possible crystallographic restrictions are imposed, it is reasonable to assume that the configuration at a threefold node will tend to be not very different from this.

One is thus led to the concept of a three-dimensional network of dislocation lines existing in a crystal. The mechanism proposed by Frank and Read (1950) for the generation of new dislocations envisages a finite length of dislocation line, 'anchored' at its ends, about which it can rotate, so producing one new dislocation loop per revolution. It has been suggested that the side of any mesh of the dislocation network may act as such a Frank–Read source. The purpose of this paper is to consider what limitations, if any, the dissociation into partial dislocations will impose on the possible dislocation nodes, and on the behaviour of the Frank–Read source. This approach draws attention to the essentially three-dimensional nature of the processes taking place, which it is easy to overlook when making a theory of work-hardening, or creep, or similar phenomena.

§2. Twofold Nodes

Throughout the following we consider that all dislocations lie in {111} planes, in which they dissociate in the manner proposed by Heidenreich and Shockley. We begin by considering the meeting of two dislocation lines. The Burgers vectors of the two perfect dislocations must be equal and opposite and, in the above notation, can be taken as, say, **AB** and **BA**. Two possibilities now arise:

(i) *The two dislocations lie in the same plane, say* (*d*). The first will have dissociated into **A**$\boldsymbol{\delta}$ and $\boldsymbol{\delta}$**B** and the second into **B**$\boldsymbol{\delta}$ and $\boldsymbol{\delta}$**A**. According to the rule given in the Appendix, the only possible arrangement of these is shown in fig. 2 (*a*), in which the lines represent the position of the dislocations, and the symbol attached to each specifies its Burgers vector, defined by an observer situated at the node O. Owing to the mutual repulsion of the partial dislocations, the configuration of fig. 2 (*a*) will be able to reduce its energy by changing into 2 (*b*), and this will therefore be the condition actually found. No restriction is placed on the directions of the partial dislocation lines in the plane. The above is, in fact, simply a description in terms of present concepts of a bend in a dislocation ribbon.

---

* The force of much of the discussion which follows depends on a clear appreciation of the geometrical relations involved. Since there is a certain element of arbitrariness in the formal definition of a Burgers vector, the viewpoint adopted here is summarized in the Appendix.

(ii) *The two dislocation lines do not lie in the same plane.* Since the Burgers vectors are **AB** and **BA**, the only possible planes in which they can lie are (c) and (d). If **AB** lies in the plane (c), say, its partials will be **Aγ** and **γB**. **BA** must then lie in plane (d), and its partials will be **Bδ** and **δA**. It is not now possible by any combination of these quantities to form two twofold nodes of partial dislocations, as was done above, and still satisfy the fundamental condition of zero vector sum at each. We can, however, satisfy this condition if we form two such nodes and then join them by another dislocation line, whose Burgers vector lies along the direction $\delta\gamma$ (fig. 3 (*a*)). This would be described as a dislocation with vector **γδ** if seen from O′ and as **δγ** when seen from O. The direction of the vector is along the join of the mid-points of two faces of the tetrahedron and its magnitude is $b/3$, as can readily be seen from the geometry of the tetrahedron (b is the Burgers vector of a perfect dislocation, i.e. the length of an edge of the tetrahedron).

The dislocation with vector **δγ** must, by definition, lie along the line of intersection of the planes (c) and (d), i.e. it must be straight and parallel to AB. Since AB and $\delta\gamma$ are necessarily at right angles, it must have edge character. Any movement, however, is dictated by possible movement of the dislocations with vectors **AB** and **BA**. These can move each in its own glide plane, and will probably be able to carry **δγ** with them; if this happens, it will be moving parallel to its own length. It is peculiar also in that it cannot exist in a perfect lattice. It represents, in fact, the singularity where a ribbon of stacking fault bends from one $\{111\}$ plane to another.* There are twelve possible Burgers vectors for such dislocations, comprising six oppositely directed pairs.

Such a stair-rod dislocation will be characterized by a certain energy per unit length. An ordinary dislocation ribbon will adopt that separation of the two partial dislocations which makes its energy per unit length a minimum. The possibility therefore arises that the system of fig. 3 (*a*) may reduce its energy by adopting the configuration of fig. 3 (*b*). The increase in energy caused by reducing the width of the ribbons at their ends has been more than compensated by the reduction in length of **γδ**. In the limit, if **γδ** were to shrink to zero length, we would have a fourfold node of partial dislocations. This is the situation which has hitherto always been assumed to exist, but the above argument shows that it is, in fact, only a limiting case, and unlikely to be found in practice.

If the signs of the quantities involved in the above argument be considered in detail, it will be found that it has been tacitly assumed that the dislocation ribbon bends from one $\{111\}$ plane to another in such a manner that the two arms make an acute angle with one another. It can also make the same change in such a manner that the included angle is obtuse, and the same kind of argument can be applied to this case also. It is found that the Burgers vector to be associated with the singularity at the bend now has a direction parallel to the line joining the mid-points of two opposite edges of the tetrahedron, and a magnitude equal to two-thirds of this length, i.e. is equal to $\sqrt{2}b/3$. There are six such vectors, three and their negatives. In the particular case shown in fig. 3 (*c*), for example, it is equal to $\boldsymbol{\gamma}\mathbf{A}+\boldsymbol{\delta}\mathbf{B}$ when seen from O, and to $\mathbf{A}\boldsymbol{\delta}+\mathbf{B}\boldsymbol{\gamma}$ when seen from O′. The sum of these two quantities $(=\boldsymbol{\gamma}\mathbf{A}+\mathbf{A}\boldsymbol{\delta}+\boldsymbol{\delta}\mathbf{B}+\mathbf{B}\boldsymbol{\gamma})$ is clearly zero as required. The direction of either is along the join of the mid-points of AB and CD. No simple notation is obvious to describe such a vector: but one which is self-consistent and tractable emerges if we note that the mid-points of AB, $\gamma\delta$, $\alpha\beta$ and CD are collinear and equally spaced. The vector in question can thus be represented by $\mathbf{CD}|\boldsymbol{\alpha\beta}$ (or $\boldsymbol{\alpha\beta}|\boldsymbol{\gamma\delta}$ or $\boldsymbol{\gamma\delta}|\mathbf{AB}$), where the order of the letters on either side of the vertical bar is immaterial (i.e. $\mathbf{CD}|\boldsymbol{\alpha\beta}\equiv\mathbf{DC}|\boldsymbol{\alpha\beta}$ etc.) and the symbol denotes a vector equal to twice the join of the mid-points of CD and $\alpha\beta$. We thus have

---

* The descriptive name 'stair-rod' dislocation has been suggested by Mr. F. R. N. Nabarro.

$$\begin{aligned}\mathbf{CD}|\boldsymbol{\alpha\beta}\ \text{etc.}&=\mathbf{C}\boldsymbol{\alpha}+\mathbf{D}\boldsymbol{\beta}=\mathbf{C}\boldsymbol{\beta}+\mathbf{D}\boldsymbol{\alpha}\\&=\boldsymbol{\gamma}\mathbf{A}+\boldsymbol{\delta}\mathbf{B}=\boldsymbol{\gamma}\mathbf{B}+\boldsymbol{\delta}\mathbf{A}.\end{aligned}$$

The negative of this vector will be denoted by $\boldsymbol{\alpha\beta}|\mathbf{CD}$ etc. Since $\mathbf{CD}|\boldsymbol{\alpha\beta}$ is perpendicular to $\mathbf{AB}$, the singularity again has edge character. We may note also that $\mathbf{CD}|\boldsymbol{\alpha\beta}$ is perpendicular to $\boldsymbol{\gamma\delta}$, i.e. the Burgers vectors of the two singularities which can lie along AB (when a ribbon bends from plane (c) to plane (d)) are at right angles to one another.

§ 3. THREEFOLD NODES

We turn now to the consideration of threefold nodes, and again there are a number of different possibilities.

(i) *All three dislocation lines lie in the same plane, say* (*d*). The principle of zero vector sum requires that their Burgers vectors be $\mathbf{AB}$, $\mathbf{BC}$ and $\mathbf{CA}$ (or the negatives of these quantities); these will split up into partials $\mathbf{A}\boldsymbol{\delta}+\boldsymbol{\delta}\mathbf{B}$, etc. There are two cases to be considered, depending on the cyclic arrangement of the three dislocation lines. The first is shown in fig. 4 (*a*). When the dislocations split into partials, these can cancel in pairs near the node and produce the configuration shown in fig. 4 (*b*). The second is shown in fig. 4 (*c*). When the dissociation takes place there is no obvious way in which they can rearrange themselves which does not increase the energy of the system. Thus there appear to be two different types of co-planar threefold node (figs. 4 (*b*) and 4 (*d*)) which can be conveniently distinguished as 'extended' and 'contracted' types respectively. Both types can move as a unit in the glide plane without difficulty.

(ii) *Two only of the dislocation lines lie in the same plane*. Let the plane be (d) and the two Burgers vectors concerned be $\mathbf{AB}$ and $\mathbf{BC}$. The third dislocation must therefore have vector $\mathbf{CA}$, and since, by hypothesis, it does not lie in (d), it must lie in (b). The partial dislocations are thus $\mathbf{A}\boldsymbol{\delta}+\boldsymbol{\delta}\mathbf{B}$, $\mathbf{B}\boldsymbol{\delta}+\boldsymbol{\delta}\mathbf{C}$ and $\mathbf{C}\boldsymbol{\beta}+\boldsymbol{\beta}\mathbf{A}$. Possible configurations can perhaps be most simply discussed by regarding them as combinations of simpler units already described. Thus we may take either of the co-planar threefold nodes (figs. 4 (*b*) and 4 (*d*)) and cause one of the ribbons to bend through either an acute or an obtuse angle, the bending taking place at the node itself. In this way the four arrangements shown in figs. 5 (*a*)–(*d*) are obtained. It is possible for all of these to move by

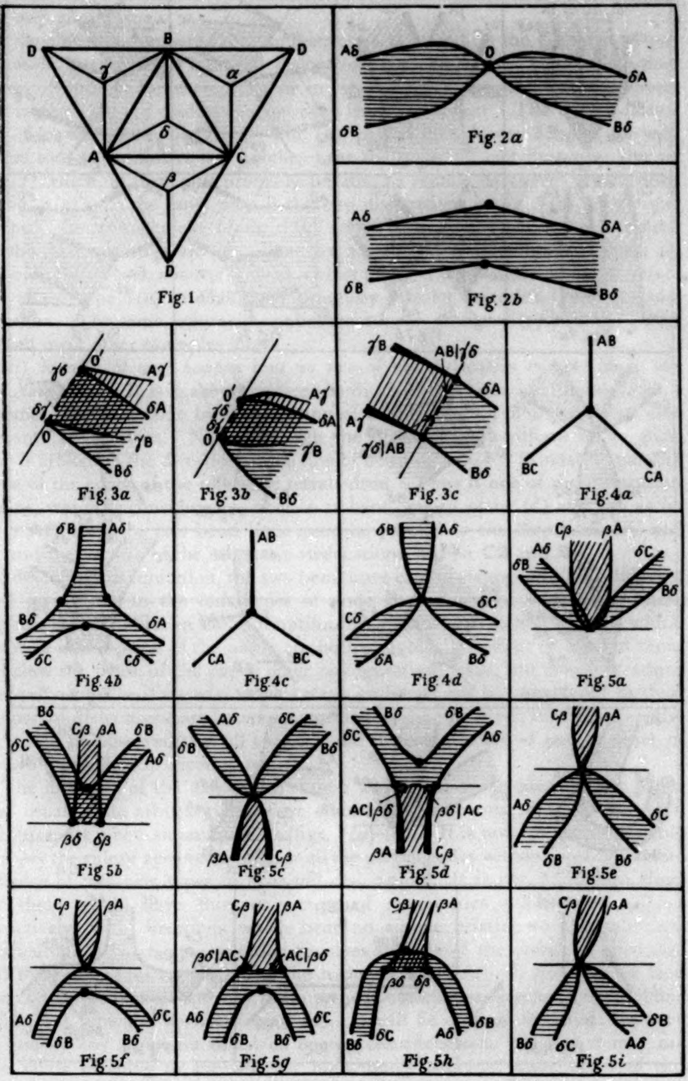

D
B
D
γ
α
δ
A
C
β
D
Aδ
O
δA
δB
Bδ
Aδ
δA
δB
Bδ
γδ
O'
Aγ
δγ
δA
O
γB
Bδ
O'
Aγ
γδ
δγ
δA
O
γB
Bδ
γB
AB|γδ
δA
Aγ
γδ|AB
Bδ
AB
BC
CA
δB
Aδ
Bδ
δA
δC
Cδ
AB
CA
BC
δB
Aδ
δC
Cδ
δA
Bδ
Cβ
βA
Aδ
δC
δB
Bδ
Bδ
Cβ
βA
δB
δC
Aδ
βδ
δβ
Aδ
δC
δB
Bδ
βA
Cβ
Bδ
δB
δC
Aδ
AC|βδ
βδ|AC
βA
Cβ
Cβ
βA
Aδ
δC
δB
Bδ
Cβ
βA
Aδ
δC
δB
Bδ
Cβ
βA
βδ|AC
AC|βδ
Aδ
δC
δB
Bδ
Cβ
βA
βδ
δβ
Bδ
δC
Aδ
δB
Cβ
βA
Bδ
δC
Aδ
δB

Fig. 1

Fig. 2*a*

Fig. 2*b*

Fig. 3*a*

Fig. 3*b*

Fig. 3*c*

Fig. 4*a*

Fig. 4*b*

Fig. 4*c*

Fig. 4*d*

Fig. 5*a*

Fig. 5*b*

Fig. 5*c*

Fig. 5*d*

Fig. 5*e*

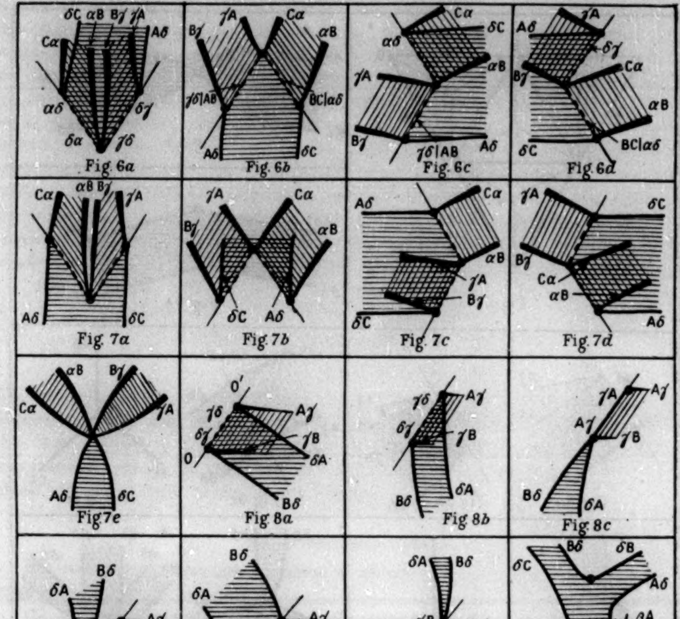
Fig. 5*f*

Fig. 5*g*

Fig. 5*h*

Fig. 5*i*

Figs. 1–5.


δC
αB
Bγ
γA
Aδ
Cα
αδ
δγ
δα
γδ
γδ|AB
BC|αδ
δC
O′
O
Aγ
γB
δA
Bδ
δB
Cβ
βA
βδ
δβ
AC(d)
CB(a)
BA(c)
P
Q
R
BC
CB
BA
AB
AB|γδ
δB
Fig. 6*a* Fig. 6*b* Fig. 6*c* Fig. 6*d*
Fig. 7*a* Fig. 7*b* Fig. 7*c* Fig. 7*d*
Fig. 7*e* Fig. 8*a* Fig. 8*b* Fig. 8*c*
Fig. 8*d* Fig. 8*e* Fig. 8*f* Fig. 9*a*
Fig. 9*b* Fig. 9*c* Fig. 9*d* Fig. 10*a*
Fig. 10*b* Fig. 10*c* Fig. 10*d*

Figs. 6–10.

glide processes but only in the direction AC, each dislocation ribbon moving in its own glide plane.

It should be remembered that no restriction is placed on the direction of the dislocation lines (except stair-rod dislocations). This brings to light another feature. Suppose, for example, that in fig. 5 (*a*) the two dislocation ribbons lying in plane (d) are made to rotate away from each other. The configuration of fig. 5 (*e*) will eventually arise. The two partial dislocations **δB** and **Bδ** will now be able to annihilate one another near the point O, and thus give rise to fig. 5 (*f*) which, in turn, will probably be able to reduce its energy by turning into fig. 5 (*g*). In the same way, if the two dislocations of fig. 5 (*b*) are caused to rotate, the arrangement of fig. 5 (*h*) would arise, and this in turn would probably contract to fig. 5 (*i*). Thus by a rotation of dislocation ribbons in their own glide plane, about a node as a pivot point, it is possible for a contracted node of the type 5 (*a*) to change continuously into an extended type 5 (*b*) and vice versa. The same argument applies, of course, to figs. 5 (*c*) and 5 (*d*), and we shall meet other examples later.

(iii) *Finally, it may happen that no two of the dislocation ribbons lie in the same glide plane.* It is again convenient to think of this situation arising from a co-planar threefold node by the bending of two (or three) of the ribbons into different {111} planes. Now, although the direction of a ribbon in its glide plane is arbitrary, the direction of the line of bending is not : it must be parallel to one of the edges of the reference tetrahedron. Thus if one of the dislocation ribbons, which is considered to remain unbent, lies in plane (d) and has as its vector **AC**, then the two bend lines must be parallel to the directions BC and BA, and the vectors of the other two dislocations will be **CB** and **BA**.

Subject to this limitation, the two bend lines can be arranged in four different ways, giving rise to the four types of node shown in figs. 6 (*a*)–(*d*). These diagrams are all drawn on the assumption that the two dislocation ribbons which do not lie in the plane of the paper are both above it. If either or both of them are below the plane of the paper other configurations arise, but these introduce no novelty into the discussion at this stage, and need not be considered further. The partial dislocations are lettered according to the rule given in the Appendix, and it will be observed that all the stair-rod dislocations are of one or other of the two types already encountered.

The direction of the dislocation ribbon which lies in the plane of the paper is, as usual, quite arbitrary and if we consider it to be rotated about the node by 180°, quite a new situation arises (figs. 7 (*a*)–(*d*)). It is now no longer possible to satisfy the rule of zero vector sum at all the partial nodes without postulating the existence of two new types of stair-rod dislocation. It is not difficult to show that these would have Burgers vectors of magnitudes $\sqrt{5}\,b/3$ and $\sqrt{7}\,b/3$ respectively, with directions which bear no simple relation to the reference tetrahedron. (Their representation involves the use of the points of trisection of the edges.) This suggests that the nodes of figs. 7 (*a*)–(*d*) may not, in fact, exist in that form at all but may take up the contracted arrangement exemplified in fig. 7 (*e*), which is derived from 7 (*a*). It will be difficult to decide between the two by any argument based on energy considerations. In fact it may not have much meaning to ask whether high-energy stair-rod dislocations of this type exist for a short distance close to a node. The lattice will be rather violently distorted for some little distance from such a point, and it may be a deceptive simplification to represent a dislocation by a geometrical line and, *a fortiori*, a contracted node by a point.*

(iv) Whatever the detailed configuration at a node may be, we can draw some general conclusions from the previous discussion. (*a*) Twofold nodes can always move by glide processes, albeit in one direction only relative to the lattice. (*b*) Threefold nodes of the type discussed under (ii) above are subject to the same limitations as twofold nodes. (c) Threefold nodes discussed under (iii) above cannot move at all by glide processes. These latter will form ideal anchor points for the operation of a Frank–Read source, but the other types are not completely excluded. (*d*) The ribbons of stacking fault will probably either all come to a point at the node (contracted type) or will all be bounded at the node by one (or two) dislocations of the stair-rod types already discussed. (*e*) Fourfold nodes, and those of higher order, are unlikely to be important, since they will usually be able to split up into threefold nodes joined by short lengths of new dislocation lines.

§ 4. OPERATION OF A FRANK–READ SOURCE

We proceed now to consider the implications of the structure of a node on the operation of a Frank–Read source. The essential feature of this process is that a dislocation rotates in its glide plane about a point on its length and gives rise to one unit of slip each time it completes one revolution. It is suggested that one possible type of anchor point for this rotation is a dislocation node. We enquire first whether the structure of the various types of node will have any effect on the first revolution, and consider later the results of repeated operation.

(i) *The First Revolution*

The behaviour is found to depend on the nature of those other singularities, in addition to the dislocation ribbon whose motion is being considered, which also lie in the glide plane. We consider four cases in turn. (i*a*) The glide plane contains no other singularity, e.g. the dislocation **CA** of fig. 5 (*a*) rotates in plane (b). The rotation proceeds unhindered and calls for no comment. (i*b*) The rotating dislocation terminates in a stair-rod, e.g. **BA** of fig. 3 (*a*) rotates in plane (d) or CA of fig. 5 (*b*) rotates in plane (b). Taking the former as an example, figs. 8 (*a*)–(*f*) represent successive stages in the motion of **BA**. As the applied stress is increased, a time will arrive when the partial dislocation **δA** lies very close and almost parallel to **$\gamma\delta$** (fig. 8 (*b*)). The resultant **γA** formed by the coalescence of these two will almost certainly have a lower energy than the two separately, and thus they will combine to give fig. 8 (*c*). The dislocation line OO′ is now an ordinary partial, and no longer constrained to lie in any particular direction; thus it may move in plane (c)—depending on the shear stresses acting in this plane—and so transform the node at O into a contracted type, with four partial dislocations meeting at a point. But whether or not this takes place, a further rotation of the moving dislocation can give rise to fig. 8 (*e*), either through the intermediate state of fig. 8 (*d*), or directly by a splitting up of the contracted node at O. The full dislocation with vector **BA** has thus pivoted round through 180°, so that whereas it started by making an acute angle with the rest of the node (stair-rod dislocation **$\delta\gamma$**), it now makes an obtuse angle

---

* We may note that similar arguments can be applied to contracted nodes of types already considered, e.g. figs. 5 (*a*) and (*c*), and even fig. 4 (*d*). It is usually possible to think of the partial dislocations uniting to form some new type of high energy for a short distance before they reach the central point.

(stair-rod dislocation **AB | CD**). If the process is continued on similar lines, we pass through the state shown in fig. 8 (*f*) and eventually return to the original configuration of fig. 8 (*a*). (i *c*) The plane of the rotation includes two stair-rod dislocations: e.g. **AC** of fig. 6 (*a*) rotates in plane (d). We have already seen that when the rotation has proceeded through 180°, the node will probably have changed from the extended type of fig. 6 (*a*) to the contracted type of fig. 7 (*e*). It is not clear what the intermediate stages will be in detail, but they will almost certainly involve a gradual shortening of the two stair-rods. A further rotation of 180° will restore the *status quo ante*. (i *d*) The plane of the rotation includes a second dislocation ribbon: e.g. AB of fig. 5 (*b*) rotates in plane (d). Here we must distinguish between the two directions of rotation. If this is anticlockwise, the two dislocation ribbons will eventually coalesce in the immediate vicinity of the node, and any further movement will come about by the branch **AB** sliding along the branch **BC**, the resulting co-planar threefold node being of the extended type (fig. 9 (*a*)). The details of a clockwise rotation are more difficult to follow. A reasonable guess would be that the rotation converted the node from the extended to the contracted type (figs. 9 (*b*)–(*d*)). **AB** would then be able to slide along the other side of **BC**, the co-planar threefold node being this time of the contracted type.

Thus in three of the four cases the splitting of the dislocations into partials gives rise to little hindrance to the rotation of one of them in its glide plane about the node; the second and subsequent revolutions will be no more difficult than the first. Moreover, these three include the important case in which all the dislocation ribbons lie in different glide planes, and the node itself is thus unable to move by glide processes. In the fourth case ((i *d*) above) only a limited rotation is possible—less than one complete revolution—and it is impossible to discuss the further behaviour without a knowledge of the nature of the adjoining nodes in the dislocation network.

(ii) *Repeated Operation of a Source*

It has been implied above that when the dislocation ribbon can complete one revolution around the node, then the *status quo ante* will have been restored, and the way be open for another similar revolution. This, however, cannot be exactly true. For when a number of revolutions have been completed, giving rise, let us say, to a number of dislocation loops from a Frank–Read source, then one half of the crystal in the neighbourhood of the source will have moved relative to the other half (separated from it by the glide plane) by an amount which is the product of the Burgers vector and the number of loops generated. This motion will, of itself, have altered the situation at the node.

Suppose, for definiteness, that we consider the node to be of that type in which the three dislocations move in different glide planes; this is the most firmly anchored of all types. Let the source be in plane (d) and have Burgers vector **AC**. Then the vectors of the other two dislocations will be either (*a*) **CB** and **BA** or (*b*) **CD** and **DA**. Consider the former possibility only for the moment. A dislocation with vector **CB** must lie either in plane (a) or in plane (d). By hypothesis it does not lie in (d); thus it must lie in (a). Similarly the third dislocation **BA** must lie in (c). Then, representing the glide plane (d) as the plane of the paper, it is probable that of the two dislocations which do not lie in (d), one will be above it and one below it. This is by no means a firm crystallographic requirement, but if we recall that dislocation lines have some of the properties of elastic cords, and that at high temperatures they will possess a considerable freedom of movement, then it will be seen that it represents a position of much greater stability than that in which both lines lie on the same side of the glide plane. If this condition is in fact satisfied, then after a number of dislocation loops have been generated, one half of the crystal, carrying with it the dislocation **CB**, will have moved as a more or less rigid body relative to the other half, which contains the dislocation **BA**.

A number of different possibilities arise, but it will suffice to consider only one of them in detail as an example. Nor is it necessary, at this stage, to consider how the full dislocations split up into partials. The initial configuration might, for example, be that shown in fig. 10(*a*); this represents a dislocation of edge type, with vector **AC**, which can act as a source in the plane of the paper, being anchored by a dislocation **CB** lying above the plane and one **BA** lying below the plane. In the glide motion both the two last named can be regarded, in the first place, as moving rigidly with the crystal in which they are situated, so that, after n loops have been produced, they will meet the glide plane in the points P and R respectively where PR lies parallel to AC and is of length nb. The source dislocation **AC** will continue to lie in the same glide plane, so that the situation in fig. 10(*b*) will arise. To satisfy the principle of zero vector sum, however, the points PQR must clearly be joined by dislocation lines with vectors as shown in fig. 10(*c*). PQR have been drawn collinear, but it will be seen at once that this need not necessarily be. Since these dislocations must lie in the plane of the paper—the glide plane—they will split up into partials as shown in fig. 10(*d*). If this state of affairs did arise, or any of the numerous possible variants of it, the source would still be able to operate as usual. The dislocation in the glide plane can slide along PR, pivot around either of the end points P and R, and slide back again on the other side of PR by processes which have already been considered.

But it is, to say the least, doubtful that the situation of fig. 10(*d*) will in fact occur: for it should be noted that all of the nodes at P, Q and R can now move by simple glide processes in the glide plane. Whether or not they will do so, and in what direction, will depend partly on the direction of the dislocation lines which do not lie in the glide plane and, to an even greater extent, on the nature of the forces exerted on these dislocations by the applied stresses. We can imagine two extremes: (i) the points P, Q and R separate to considerable distances once the applied stress has done sufficient work to break up the original threefold node. (ii) The points P, Q and R continually move together again as each new dislocation loop is formed. This is the more interesting possibility, and has the effect of moving the anchor point of the source a distance b (the Burgers vector) in a direction at right angles to **b** for every loop formed. Moreover there need be no co-ordination between the behaviour of the anchor points at the two ends of any one source, so that the source length may either increase or decrease as successive loops are formed. Alternatively, a dislocation originally of purely screw character may develop some edge character as the glide proceeds.

The second possibility mentioned at the beginning of this section (namely that the two anchoring dislocations have vectors **CD** and **DA**) can be discussed in a similar manner. Since **CD** and **DA** both have a component normal to the glide plane (d), then, if they lie on opposite sides of it, the sweeping dislocation, on completing each revolution, will have moved out of its glide plane by one lattice spacing. This gives rise to a number of interesting possibilities, consideration of which, however, must be left to another occasion.

§ 5. CONCLUSIONS

The main results of this discussion may be summarized as follows:

(*a*) It appears possible for a dislocation ribbon in a face-centred cubic lattice to bend from one {111} plane to another without the two partial dislocations coming together at the bend. As a corollary, the same kind of behaviour may be expected at a 'jog' in a dislocation (see Mott 1951).

(*b*) Of the various threefold nodes which can be imagined, some can move by glide processes, but only in certain restricted directions relative to the crystal lattice, while others are unable to move by glide processes at all.

(*c*) Such threefold nodes, and particularly the latter group, may act as the anchor points at the ends of a Frank–Read source of dislocations. In this event the structure of the node will not usually prevent the operation of the source, but it is not certain that it may not provide a certain energy barrier which has to be surmounted each time a dislocation loop is generated.

(*d*) With the repeated operation of a Frank–Read source, even the most firmly anchored node may progressively move. In this way it is possible that such a source may necessarily change its length (in either sense) as the production of new dislocation loops proceeds.

In default of any quantitative estimates of the relative energies of the various arrangements of dislocations, much of the discussion must remain somewhat speculative, and it is doubtful if any useful purpose would be served by any more detailed study of the particular questions here considered. It is possible, however, to use with advantage the notation of the present paper to discuss other problems of dislocation geometry, such as the formation of sessile dislocations. The 'dislocation reactions' discussed by Cottrell (1952) can be described as follows: If a dislocation with vector $\mathbf{DC}$ lying in plane (b) and another $\mathbf{CB}$, lying in plane (d), both happen to lie parallel to AC, then they can combine to give a dislocation with vector $\mathbf{DB}$, whose line must also lie parallel to AC, i.e.

$$\mathbf{DC}\,(\mathrm{b})+\mathbf{CB}\,(\mathrm{d})=\mathbf{DB}\,(\mathrm{b},\,\mathrm{d}),$$

where the notation of the right-hand side indicates immediately that the dislocation lies along the intersection of planes (b) and (d) and that it cannot glide in a {111} plane. This is the equivalent of

$$\tfrac{1}{2}a[10\bar{1}]+\tfrac{1}{2}a[011]=\tfrac{1}{2}a[110].$$

If the original dislocations are both dissociated before the reaction takes place, we have instead

$$\boldsymbol{D\beta}+\boldsymbol{\beta C}+\boldsymbol{C\delta}+\boldsymbol{\delta B}=\boldsymbol{D\beta}+\boldsymbol{\beta\delta}+\boldsymbol{\delta B},$$

where it is no longer necessary to specify the planes explicitly. Here the presence of the component $\boldsymbol{\beta\delta}$ emphasizes the fact that the dislocation lines must have a certain direction (i.e. parallel to AC) and that the resulting combination is sessile in the restricted sense of Lomer and Cottrell. The equivalent reaction in the other notation is

$$\tfrac{1}{6}a[11\bar{2}]+\tfrac{1}{6}a[2\bar{1}\bar{1}]+\tfrac{1}{6}a[112]+\tfrac{1}{6}a[\bar{1}21]=\tfrac{1}{6}a[11\bar{2}]+\tfrac{1}{6}a[112]+\tfrac{1}{6}a[110].$$

ACKNOWLEDGMENTS

It is a pleasure to acknowledge the assistance derived from discussions with numerous colleagues, particularly Dr. F. C. Frank, during the development of these ideas.

APPENDIX

The formal treatment here given is based on that of Frank (1951). We first define the Burgers vector of a dislocation line. We choose any point on the line as origin and assign, arbitrarily, the positive sign to one direction of the line leaving this origin. Looking outwards from the origin along the positive direction of the line, we make a circuit in the crystal in a *clockwise* sense: this circuit must link the dislocation line and close on itself. We make a 'corresponding circuit' (Frank 1951) in a similar and similarly orientated perfect crystal, i.e. one containing no dislocation line. This circuit will not close on itself. The additional displacement needed to close the circuit in the 'reference lattice' is the Burgers vector of the original dislocation.

We next give a rule for determining the direction of the slip movements involved when a specified dislocation line moves in a specified manner. We consider an observer to be situated in either of the semi-infinite crystals separated by the glide plane, and to be looking towards this plane. We make a rotation of 90° in a *clockwise* sense from the positive direction of the dislocation line; this defines the positive direction of dislocation movement. Then when the dislocation line passes across the glide plane in the positive direction as so defined, that half of the crystal towards which the observer is looking will move relative to the half in which he is situated by an amount equal to, and in the direction of, the Burgers vectors.

Thirdly, we consider the question of partial dislocations in face-centred cubic crystals. As has already been stated, it is convenient to refer to a regular tetrahedron, whose faces are parallel to the four {111} planes of the lattice. It should be noted that there are two essentially different ways of constructing and lettering such a tetrahedron, which are mirror images of one another. One of these should be chosen and used consistently. Once chosen, it is uniquely orientated relative to the crystal. All operations are to be carried out from the viewpoint of an observer situated *outside* the tetrahedron.

Consider now a dislocation with Burgers vector, say **BC**, and lying in plane (d). If it splits up into two partial dislocations their vectors will be **Bδ** and **δC**, but we must now set up a rule to decide which vector is to be associated with which dislocation line. This is as follows: The positive direction of the dislocation line is already specified; then for an observer situated outside the tetrahedron, and looking in the positive direction, the right-hand component will be **Bδ** (Roman–Greek) and the left-hand component **δC** (Greek–Roman).

If the rule given above for the direction of slip is applied to the passage of such a dissociated dislocation across a glide plane, it will be found that the sequence of movement of the atoms so specified is consistent with the zigzag movement which would clearly take place. Moreover, in a perfect crystal no other mode of dissociation is possible.

REFERENCES

COTTRELL, A. H., 1952, *Phil. Mag.*, **43**, 645.

FRANK, F. C., 1951, *Phil. Mag.*, **42**, 809.

FRANK, F. C., and READ, W. T., 1950, *Phys. Rev.*, **79**, 722.

HEIDENREICH, R. D., and SHOCKLEY, W., 1948, *Bristol Conference on Strength of Solids* (London : The Physical Society), p. 57.

MOTT, N. F., 1951, *Proc. Phys. Soc.* B, **64**, 729.

A Simple Varying Capacitor Method for the Measurement of Contact Potential Difference in High Vacuum

By H. P. MYERS

Department of Natural Philosophy, Marischal College, Aberdeen

MS. received 23rd January 1953

Abstract. Methods of measuring work function are discussed. An apparatus based upon Kelvin's original method for the measurement of contact potential difference and suitable for use in vacua of 10^{-8} mm Hg is described. The value of contact potential difference between copper and silver films evaporated on to tungsten sheets was found to be $0{\cdot}28 \pm 0{\cdot}03$ volt, the silver being positive with respect to copper.

§ 1. INTRODUCTION

THE principal methods for the determination of the work function of a metal which depend upon either the study of thermionic emission or its photoelectric properties have certain limitations. Accurate thermionic emission data can be obtained only with the highly refractory metals such as tungsten, molybdenum or tantalum, and this method has little application to the lower melting point metals. Methods based upon the photoelectric effect have general applicability but require difficult and elaborate techniques if the best results are to be obtained; this limits their use as methods available to all laboratories. The patchy or non-uniform structure of some surfaces may also result in the determination of values of work function which are not representative of the surface by these methods.

The contact potential difference between two clean metal surfaces is numerically equal to their difference of work function; this fact enables the work function to be determined from contact potential difference measurements. Such measurements have usually involved the use of free electrons as in the diode, electron beam and magnetron methods. The use of electron beams is a disadvantage when measurements are made on patchy surfaces.

It is difficult to see how patchy surfaces can be avoided, for the studies of thermionic emission, field emission and contact potential difference measurements on single crystals provide reliable evidence for the variation of work function with crystal face (Herring and Nichols 1949). In general, therefore, a polycrystalline film of metal presents a surface comprised of patches of different work function; these patches may differ in work function to the extent of several tenths of a volt, and the value measured for the whole surface may depend upon the method of measurement used. Measurements based on the photoelectric effect yield values weighted in favour of the areas of low work function. Contact potential difference measurements with the electron beam method as used by Anderson (1949) and Mitchell and Mitchell (1951), involving the comparison of (electron current, retarding potential) characteristics for a beam of electrons striking the surfaces, would also be affected by the patchy nature of the surfaces, although not to the same extent as the photoelectric method. A further consideration which limits the use of the electron gun method is the effect of reflection of the electrons by the surfaces under investigation. If these surfaces possess different electron reflection ratios for electrons with thermal energies, then the parallelism of the cut-off characteristics can be changed; furthermore, the relative displacement of these curves is also altered, resulting in errors in the measured contact potential difference. This method should therefore be used only when the two substances have similar electron reflection ratios for thermal electrons; the latter condition is satisfied when the retarding potential characteristics have identical shapes in the region immediately preceding and during cut-off.

The original method of Kelvin does not encounter the above difficulties; the true average contact potential difference between the surfaces is measured, and the method lends itself to the investigation of processes affecting the work function of a metal more easily and more reliably than the former methods. The metals whose contact potential difference is required form the active surfaces of a condenser. If a contact potential difference exists and the capacity of the condenser is changed, a flow of charge results. When one plate is caused to vibrate, an oscillatory flow of charge is obtained which can produce a signal voltage capable of being amplified and registered on a cathode-ray oscilloscope. This signal may be reduced to zero by applying an external potential in series with the condenser; this potential is then equal and opposite to the contact potential difference.

This principle has been used in the design of apparatus for the measurement of contact potential difference in air and in vacua of 10^{-5} mm Hg, but such apparatus is not suitable for operation under the best vacuum conditions. The method described below may be used under the highest vacuum conditions, and is easily adapted for studying factors affecting the work function of metals. In this apparatus a varying capacitance is produced by an extremely light mechanical shock causing the vibration of a condenser plate attached to a tungsten wire.

§2. APPARATUS

The condenser was formed by two circular plates of tungsten sheet, diameter 2 cm and thickness 0·0025 in. One of these plates was spot welded to a straight tungsten wire, diameter 0·3 mm, length 2·5 cm, the other end of the wire being welded to a nickel rod, diameter 1 mm. The second plate carried a polished tungsten wire, original diameter 0·6 mm, polished to a diameter 0·55 mm, mounted at right angles to the plane of the disc. This wire was polished electrolytically until it would just slide easily into a short length of stainless steel tubing, length 6 mm, this tube being attached by nickel tape to a nickel rod, diameter 1 mm. The plate was prevented from sliding out of the tube by welding a small piece of nickel rod to the end of the tungsten wire to act as a stop (fig. 1). The polished wire mount for the plate moved freely in the tube, which formed a convenient guide and support for the second disc.

These plates were attached to the two central wires of a ten-wire pinch. The nickel rods were bent and cut so that the plates were parallel, opposite one another, and separated by a gap of not more than 1 mm when the nickel stop was against the steel tube. On tilting the pinch the plates could then be separated by a distance of about 2 cm by allowing the second plate to slide in the guide tube. In this way the plates could be separated or brought close together merely by a slight tilt of axis.

The remaining eight wires of the pinch were distributed about its circumference and served to carry the screen, evaporators and other accessories described below. A factor to be considered in the design of this apparatus was the shielding of the condenser and supports from stray electrostatic charge that might reside on any neighbouring insulator surface: in addition, it was necessary to screen from unwanted signals. For this purpose a molybdenum screen in the form of a cylinder, length 1·5 in., diameter 1·5 in., sheet thickness 0·0025 in., with partially closed ends, was used to surround the condenser assembly. The top of the cylindrical cage was closed by two segments of molybdenum sheet, allowing sufficient room for the nickel rods to pass through. The bottom of the cage was closed in a similar fashion, but a 1 cm gap running parallel to the condenser plates was left. In this gap a helical filament of 0·2 mm diameter tungsten wire was mounted, and was used to de-gas the condenser plates by electron bombardment.

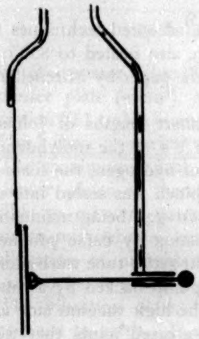


Fig. 1. The arrangement of the condenser plates.

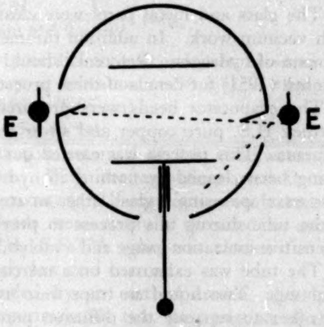


Fig. 2. Cross section through the central plane of the apparatus showing the condenser plates separated for the deposition of metal from the evaporators E.

The metals used were deposited on to the tungsten plates by evaporation *in vacuo*. For this purpose evaporators consisting of filaments of molybdenum wire, diameter 0·3 mm, were used as shown in fig. 2.

The molybdenum cage also served as a screen for the evaporators, which were mounted in the central plane of the system, one on either side of the cage. In order that the evaporated metal would strike the plates when at their maximum separation, two doors were cut into the molybdenum cage and bent inwards as shown in the diagram. Each evaporator could deposit metal at oblique incidence on one plate only; furthermore, metal from one evaporator could not strike the other.

All the parts essential to the measurement of a contact potential difference were carried by one pinch which sealed comfortably into a tube, diameter 3 in., length 7 in. Gauge and getter tubes were added in the usual way. All parts were fastened both to one another and to the pinch by spot welding, and all were capable of adequate de-gassing either by electron bombardment or by eddy current heating.

A further advantage of this design was that when the metals were evaporated almost the whole of the inside surface of the glass envelope was covered with metal. This layer acted as a getter and, furthermore, made a second Faraday cage to screen the condenser system since the conducting layer could be earthed by a tungsten wire spring contact to the molybdenum cage, which itself was always maintained at earth potential when measurements were made. The parts not covered by evaporated metal were the two central pinch wires, where extremely good insulation was required; these were screened by the molybdenum cage. A small area at the bottom of the tube and two narrow longitudinal sections did not receive evaporated metal; these parts were therefore initially covered with a layer of aquadag which was baked into the surface. It was convenient to make the connection from the cage to this aquadag covering. After evaporation the whole of the inner surface was covered with a conducting layer.

The glass and metal parts were cleaned using the accepted techniques for high vacuum work. In addition the metal parts were also heated to 850°c in a stream of hydrogen. Reference should be made to the paper by Mitchell and Mitchell (1951) for details of these procedures.

The evaporator beads were prepared by fusing short lengths of Johnson Matthey H.S. pure copper and silver wire on to the V's of the molybdenum filaments. This process was carried out in a stream of hydrogen, the filament having been cleaned by flashing in hydrogen. The pinch was sealed into the glass envelope using a glass lathe, an atmosphere of nitrogen being maintained in the tube during this process to prevent contamination by flame products. A sensitive ionization gauge and a molybdenum filament getter tube were added.

The tube was exhausted on a mercury diffusion pump backed by a rotary oil pump. Two liquid air traps were used, one on the high vacuum side and the other to separate the diffusion pump from the greased joints that were used in the backing side. This latter trap was immersed in liquid air throughout the whole of the pumping procedure.

The usual high vacuum practice was followed, and between baking at 450°c the metal parts were de-gassed. The molybdenum cage was easily heated to 1000°c by eddy current heating. The cage, however, screened the two tungsten plates which at this stage were always placed at their maximum separation; they were de-gassed by heating to 1000°c by electron bombardment. The ionization gauge was de-gassed by electron bombardment and the getter filament by flashing to above 2000°c.

Before sealing the tube from the pump the evaporator filaments were repeatedly flashed to de-gas the metal beads and sufficient metal deposited to cover the inside of the glass envelope. The constriction was de-gassed at dull red heat and the tube sealed off with both evaporators and the getter filament hot. The pressure after sealing from the pump was never greater than the limiting reading of the ionization gauge, $3\text{–}5 \times 10^{-8}$ mm Hg.

The tungsten plate always moved easily in the guide tube after the de-gassing treatment described.

Variation of the capacity of the condenser formed when the plates were at their minimum separation was provided in a very simple manner. Gentle tapping of the glass envelope caused the plate attached to the tungsten wire to vibrate; being in a vacuum the plate vibrated quite freely, the internal friction of the tungsten being the only source of damping.

§ 3. THE ELECTRICAL CIRCUIT

The condenser was placed in series with a potentiometer, the whole being shunted by an 8-megohm high stability carbon resistor. The signal generated across this resistor was first amplified by a single valve amplifier using an E.F.37A tube and then by the amplifiers of a cathode-ray oscilloscope. An estimate of the signal voltage developed across the resistor is given below. This estimate is based upon the assumption that the distance between the condenser plates varies sinusoidally with time. The exact nature of the vibration is not simple harmonic in form, but a reliable estimate may be made with this assumption.

δV, the voltage across the resistor, is given by

$$\delta V = \frac{RVA\alpha p}{4\pi d} \frac{\cos pt}{(1+\alpha \sin pt)^2} \frac{1}{9\times 10^{11}} \text{ volts,}$$

where R = resistor value ($8\times 10^6 \Omega$), V = potential across condenser, A = area of condenser plate (π cm^2), $p/2\pi$ = frequency of vibration ($\sim$50 c/s), d = normal separation (0·1 cm), α = amplitude of vibration (0·05 cm).

Using these numerical values, $\delta V_{\max} = 5\times 10^{-3}$ v.

The maximum signal for a difference in potential of 0·01 v between the condenser plates is therefore 5×10^{-5} v. With the amplifiers used it was possible to measure potentials to this accuracy. Precautions were taken to eliminate microphonic noise, and all leads were carefully screened to avoid amplification of unwanted signals. The E.F.37A tube was chosen for its anti-microphonic properties and extremely low grid current.

The vibrating plate, being surrounded by an earthed screen, also had capacitance to earth. It was shown experimentally that the variation of this capacitance caused when the plate vibrated was too small to influence the measurements.

§ 4. PROCEDURE

All electrodes other than the condenser plates were earthed. The condenser plates were separated by tilting the tube and the metals under investigation evaporated; the movable plate was then placed into the measuring position. The wall of the tube was gently tapped with the finger and the signal on the cathode-ray screen reduced to zero by adjustment of the potentiometer. A millivoltmeter measured the external potential applied to the condenser. With this apparatus a measurement of contact potential difference to within ± 0·01 volt could be made in less than one minute after the evaporation had been completed.

The contact potential difference between evaporated layers of copper and silver was found to be 0·285 ± 0·03 v, silver being positive with respect to copper. This result is the mean of measurements on ten pairs of films. In the first tube of this pattern a nickel cage was used, but nickel evaporated on to the evaporator beads during eddy current heating; the evaporators were therefore contaminated with nickel, which readily forms solid solutions with copper and silver. This tube gave a result 0·4–0·35 v, but no significance has been attached to it for the above reason. Molybdenum metal has been used in all subsequent tubes.

No change in contact potential has been found to occur with time after the evaporation: the same value of potential has been measured a week after evaporation as was measured immediately after the evaporation was made. It is considered that the values given above are representative of clean copper and silver surfaces.

The excellent vacuum conditions are attributed not only to the careful pumping schedule but also to the gettering action of the large area of evaporated metal that covered the inside of the glass envelope. It was not necessary to immerse the getter tube in liquid oxygen to maintain the vacuum.

§5. DISCUSSION

The value of contact potential difference determined in this work may be compared directly with the measurement made by Mitchell and Mitchell, who used the electron beam method; their value for the contact potential difference is 0·28 ± 0·05 volt. The agreement between the two measurements is very good. In both cases measurements were made on Johnson Matthey H.S. pure copper and silver metal, tungsten being used as a substrate and good vacuum conditions maintained.

Further discussion of Mitchell and Mitchell's value may be made in the light of recent measurements of electron reflection ratios at low energies (Myers 1952). This work showed that clean surfaces of copper and silver evaporated on to de-gassed tungsten possessed electron reflection ratios 0·1 and 0·12 respectively for electrons with thermal energies. Tungsten immediately after being heated to above 2000°c also possessed an electron reflection ratio of 0·12 for electrons with thermal energies. The similarity of electron reflection ratio for the three materials is an assurance that the reflection effect is of no significance in connection with Mitchell and Mitchell's measurements on copper, silver and tungsten. The comparison of the two measured values of contact potential difference indicates that for these metals the electron gun method gives a close approximation to the average contact potential difference between the surfaces, the values of work function determined being consistent with the contact potential difference measured by the Kelvin method.

In the light of the above discussion the values of work function determined by Mitchell and Mitchell for copper and silver layers evaporated on to tungsten may be taken as standard values for these surfaces. Copper or silver layers deposited on tungsten therefore form convenient standards of comparison for further work function measurements by the Kelvin method.

The measurement of the contact potential difference between copper and barium and silver and barium by Anderson shows the contact potential between copper and silver films to be zero. These metals were however deposited on glass substrates. There can be no question of the difference between this result and those presented in the former paragraphs being due to contamination or faults in the method of measurement, and both sets of measurements should be considered to be reliable and representative of the materials used. To explain the divergence of the two sets of results it is necessary to assume that the kind of substrate used has a pronounced effect upon the condition of the surface and its work function.

The recent measurement by Anderson (1952) of the work function of silver evaporated on to tantalum sheet (4·31 ev) is 0·16 ev less than his earlier value for silver on glass and agrees very well with Mitchell and Mitchell's value (4·33 ev). Tantalum and tungsten possess similar crystalline structures and have approximately similar lattice constants. Furthermore, in sheet form they both present coarsely crystalline surfaces; they may therefore be considered as equivalent substrates. The experiments of Anderson (1936, 1941, 1949, 1952) and Mitchell and Mitchell may therefore be taken as further evidence for the structure dependence of the work function. More direct tests could be made by measuring the contact potential difference of silver (copper) on glass against silver (copper) on tungsten using the Kelvin method.

ACKNOWLEDGMENTS

The author wishes to acknowledge the co-operation of Mr. R. Sherwood, whose skilled glass working was essential to the completion of this work.

Although this apparatus has been designed wholly since the author has been at Aberdeen, he wishes to acknowledge the experience gained whilst working with Dr. J. W. Mitchell at the H. H. Wills Physical Laboratory.

While these measurements were being made at Aberdeen, J. C. Rivière was undertaking an independent investigation of the same problem at Bristol, where the measurement of contact potential difference and work function form part of the research programme.

REFERENCES

ANDERSON, P. A., 1936, *Phys. Rev.*, **49**, 320; 1941, *Ibid.*, **59**, 1034; 1949, *Ibid.*, **76**, 388; 1952, *Ibid.*, **88**, 655.
HERRING, C., and NICHOLS, M. H., 1949, *Rev. Mod. Phys.*, **21**, 185.
MITCHELL, E. W. J., and MITCHELL, J. W., 1951, *Proc. Roy. Soc.* A, **210**, 70.
MYERS, H. P., 1952, *Proc. Roy. Soc.* A, **215**, 329.

Some Properties of Hydrogen Spark Channels

By R. D. CRAIG* and J. D. CRAGGS
Department of Electrical Engineering, University of Liverpool

MS. received 12*th May* 1952, *and in amended form* 18*th February* 1953

Abstract. Measurements of voltage gradients and visible channel diameters for hydrogen sparks (100–200 A, rectangular current pulses) at 1 atmosphere pressure, are described. The results are discussed in terms of the physical properties, such as the gas temperature, of the channel. It appears from the drift current equation, on insertion of measured values of current and electron density, that the electron drift velocity fits a constant pressure (arc-like) condition for the channel better than a constant density condition. The results are consistent with a channel temperature of about 15 000°. Diffusion processes, in the later stages of spark channel expansion, are also discussed.

---

§ 1. INTRODUCTION

THE physical processes involved in the formation, expansion and final decay of gaseous spark channels have been studied by various workers over a considerable period. In the present paper it is proposed to discuss the voltage gradients and channel expansion rates in short sparks in hydrogen of cylinder grade purity (0·5–1% oxygen) at about atmospheric pressure using rectangular current pulses of the type described in previous publications (Craggs and Hopwood 1947 a, b).

The work on short sparks, with rectangular current pulses lasting, say, 1–10 μsec has so far included for hydrogen discharges preliminary measurements of afterglow durations, etc , and determination of ion concentrations from Stark effect measurements and excitation temperatures (Craggs, Williams and Hopwood 1948). The present data on spark channel radii have been combined with the Stark profile work to extend the earlier experiments on ion concentration measurements, and values for mean ion concentrations are used in the present paper for the purpose of investigating further the electron drift current equation (see § 4).

In order to clarify the scope of this paper, it is of value to refer to fig. 1 (Plate I) which shows the development of a hydrogen spark channel at the mid-gap position with various current pulses, 1, 4 and 10 μsec in duration (110 amperes). This is reproduced schematically in fig. 2 where the regions AB, BC and CD are: AB, the initial part of the channel expansion process starting from the streamer channel at A which is not normally distinguishable in this type of record; BC, the period during which the channel expansion is relatively slow although the constant current is still flowing; CD, the afterglow period. The present paper refers mainly to BC, with a little discussion of AB. Recent work on the afterglow period will be reported elsewhere.

---

* Now at Metropolitan-Vickers Research Laboratories.

Figure 1 was taken with the streak camera technique in which a vertical image of the spark channel is focused on to a horizontal slit, arranged to pass light from a chosen part of the channel; usually, as in the present work, the centre of the channel is used to avoid the effects of metal vapour as far as possible. Then the image of the selected part of the channel is focused, after reflection from a rotating mirror, on to a photographic plate (as shown in fig. 6) and as the mirror rotates a streak image, showing radial growth as a function of time, of the central part of the channel is seen.

In the following, the application of the drift current equation $I=nev$ to the spark channel will be discussed (I is the current density, n the electron density, and e and v respectively the charge and drift velocity of the electron). The purpose of this paper is to report measurements of I, based on measured radii of the luminous channels, and X, the longitudinal electric potential gradient in the channel for discharges with simple and sharply defined current pulse shapes. From previous work n was obtained for the same discharge conditions, and so v can be deduced either from I and n or, on taking assumed values of T_e and T_g (the electron and gas temperatures in the channel), from X. It is thus possible to make certain deductions about T_g and the conditions in the channels; these are of value in the absence of direct measurements, which are difficult to make and not likely to be available for a considerable time.

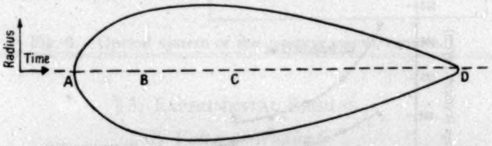


Fig. 2. Schematic representation of spark channel expansion.

In particular it has often been assumed, as mentioned below, that the spark channel expands from its narrow state, existing at the instant of complete bridging of the gap with the breakdown plasma, to its final state (~1 mm diameter for currents ~100 A) at constant density. Recent work (Drabkina 1951) on the initial shock wave expansion process (AB of fig. 2) suggested that, contrary to the above, a constant pressure state is quickly reached and evidence in support of this is given below.

The distinction between constant pressure and constant density régimes should be clearly drawn. An example of the former is an arc in equilibrium, whose properties are time independent, and which has a plasma pressure equal to that outside the discharge region. The latter is typified by the earliest part of a spark channel expansion when the shock point coincides with the luminous boundary of the channel (as shown by Gegechkori 1951) which then has a high internal pressure (of the order of tens of atmospheres). After a time of about 0·5 μsec (see fig. 1) the channel expansion rate slows down, and the shock wave separates from the luminous envelope (Gegechkori 1951) which then expands by a diffusion process, discussed below.

§ 2. EXPERIMENTAL TECHNIQUES

The current pulses for the present work were obtained with artificial lines or coaxial cables, in the normal radar modulator technique. A typical current pulse is shown in fig. 3 (Plate I).

The voltage gradients were measured in the present work with a specially manufactured cathode-ray tube in which the insulation of one pair of deflector plates was made to withstand 20 kv, the maximum voltage applied to the spark gaps, without serious alteration of sensitivity at lower voltages. Thus, the voltage oscillogram of a spark showed the spot on the oscillograph screen only after about 0·5 μsec from spark breakdown. The oscillograph screen was carefully calibrated to allow for trapezoidal distortion, which was unfortunately appreciable; typical oscillograms are shown in fig. 4 (Plate II) and some corrected voltage gradient–time records, taken from the results of tests with different gap lengths in order to eliminate the effect of the electrode drop regions are given in fig. 5. The oscillograms were taken with repeated sparks and the results showed satisfactory reproducibility. The differences in the voltage gradients measured (*a*) at the end of the 6 μsec pulses and (*b*) at the beginning of the recorded data for the 10 μsec pulses are due to slight differences in the currents obtained with the different lines. It is estimated that the experimental errors in the voltage gradients (as in fig. 5) are about ± 15% for $t = 1$ μsec and ± 5% for $t = 2$ μsec. The difference between the gap spacing and the true spark channel length was checked with photographs of the sparks and found to be negligible.

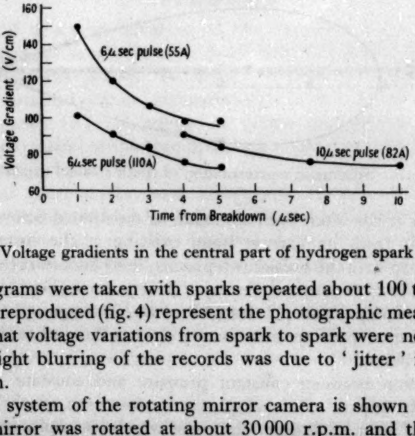


Fig. 5. Voltage gradients in the central part of hydrogen spark channels.

The oscillograms were taken with sparks repeated about 100 times per second and the records reproduced (fig. 4) represent the photographic mean of 1000 traces. It was found that voltage variations from spark to spark were negligible ($\not> 5\%$) and that the slight blurring of the records was due to ' jitter ' in the time-base synchronization.

The optical system of the rotating mirror camera is shown schematically in fig. 6. The mirror was rotated at about 30 000 r.p.m. and the single sparks necessary for the cross section streak photographs (see fig. 1) were tripped with the pulse from a multiplier photocell, illuminated at the appropriate instant by means of a small auxiliary reflector attached to the shaft of the main rotating mirror, so that close synchronism (to within about 0·1 μsec) between the correct mirror position and the spark initiation was obtained.

The writing speed of the mirror M (fig. 6) on the film P was 1 mm/μsec and the overall magnification of the channel diameter by lenses L_1, L_2 was about unity. The temporal resolution of the camera may be defined as the time taken for the image of the slit S to move across a fixed point on the film P. Slit widths corresponding to resolution times of from 0·1 to 0·5 μsec were used. Film calibrations were made by varying the overall transmission of the optical system. In the early work this was done by interposing graded gauzes at lens L_1 but it was subsequently found more satisfactory to place a Kodak neutral step filter in front of the slit S. About twelve streak photographs were taken for each gauze or step and the discharge conditions were maintained constant throughout. The mean density at a fixed point (e.g. midway across the channel and 3 μsec after breakdown) on the records was plotted against the corresponding value for the transmission of the optical system.

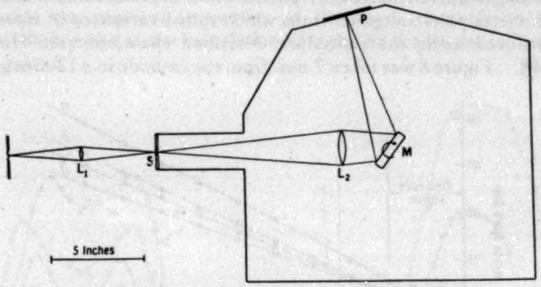


Fig. 6. Optical system of the rotating mirror camera.

§ 3. EXPERIMENTAL RESULTS

(i) *Voltage Gradients*

The voltage drop data for discharges between tungsten electrodes are shown in figs. 5 and 7. Figure 7 shows typical (gap voltage–gap spacing) curves taken for various times during a 110 A pulse of 6 μsec duration. The variation of voltage gradient in the region of the electrodes is apparent. Sufficiently far from the electrodes, the voltage gradient is constant and corresponds to that found by Higham and Meek (1950 a) who generally used higher currents. Figure 5 gives voltage gradients, taken for the constant gradient part of, for example, fig. 7, as measured for various times during the period of current flow. The voltage falls so rapidly (from 15 kv) during the first microsecond that no measurements during this time were attempted.

Detailed confirmation of the independence of channel voltage gradient (as shown in fig. 7) on the electrode material is provided by data obtained for gap lengths varying from 4 to 10 mm, lengths very much greater than the electrode drop regions, as indicated for example in fig. 7. The pulse length was 6 μsec with a current of 110 amperes.

The following electrode materials were used: Brass, Zn, Mg, Ag, Sn, Cu, Fe, Mo, Pt and W. The voltage gradients at times of 1, 2, 3 and 5 μsec for all the above were found to be respectively in the ranges 107 ± 5, 92 ± 7, 84 ± 6, and 73 ± 5 v/cm, i.e. within the range of experimental error.

(ii) *Apparent Channel Diameters*

The measurements of the apparent variations of light intensity across the channel are exemplified by the series shown in figs. 8 (*a*)–(*h*), for H_α radiation isolated with a Wratten No. 27 filter, taken with 4 μsec current pulses (110 A). Figure 8 (*a*) taken at $t = 0{\cdot}5$ μsec shows a central trough, reported also by Holtham and Prime (1950) using 1000 A pulses in this laboratory. The curves of fig. 8 refer to the microphotometer tracings of the streak photographs of the channels (such as those shown in fig. 1) corrected for the film or plate response measured as described in § 2 and are termed $I(x)/x$ plots. $I(x)$ denotes the intensity of the discharge measured external to the channel at distance x from the diameter parallel to the line of sight and perpendicular to the length of the channel; $I(x)$ is thus the integrated light, travelling in a given direction, from the whole depth of channel at x. Typical microphotometer plots are given from which radial variations of intensity $I(r)/r$ may be deduced using the procedure described elsewhere (see Holtham and Prime 1950). Figure 8 was taken 7 mm from the cathode in a 12 mm gap.

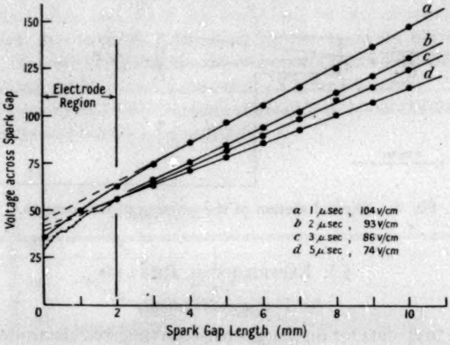


Fig. 7. Total spark gap voltages (*a*) at 1 μsec, gradient = 104 v/cm; (*b*) at 2 μsec, gradient = 93 v/cm; (*c*) at 3 μsec, gradient = 86 v/cm; (*d*) at 5 μsec, gradient = 74 v/cm.

It will be noticed, as shown clearly in the streak photographs (fig. 1), that the rate of channel expansion after about 0·5 μsec is small compared with that during this early period which shows a mean radial expansion rate of about 1 mm/μsec. When the current ceases to flow (figs. 8 (*f*) to (*h*)) the channel apparently decays in such a way as to suggest an effective contraction. Figure 9 shows all stages of the expansion process for 4 and 10 μsec discharges at 110 A (cf. fig. 2). These afterglows will be discussed in more detail elsewhere.

§ 4. DISCUSSION

The channel cross sections of the present paper, together with the analysis of line widths previously undertaken enable radial variations of ion concentration (Saxe 1948) to be made from which mean ion concentrations, measured for the luminous channel cross section, can be found. Then, using the drift current equation

$$I = nev \qquad \ldots\ldots (1)$$

which ignores the positive ion drift current, v may be determined. The data for two examples are as follows:

Case 1. $t = 3\cdot9\,\mu\text{sec}$ in 110 A, $4\,\mu\text{sec}$ discharge; $X = 75\,\text{v/cm}$, mean $n = 7\cdot5 \times 10^{16}\,\text{cm}^{-3}$ (note that this is one half of the Stark 'ion concentration' which includes electrons and positive ions, Craggs and Hopwood 1947 a), $v = 8\cdot8 \times 10^{5}\,\text{cm sec}^{-1}$.

Case 2. $t = 2\,\mu\text{sec}$ in 110 A, $4\,\mu\text{sec}$ discharge; $X = 90\,\text{v/cm}$, mean $n = 9\cdot5 \times 10^{16}\,\text{cm}^{-3}$, $v = 7\cdot0 \times 10^{5}\,\text{cm sec}^{-1}$.

Then, if the electron mobility as a function of X/p, T_g (the gas temperature) and the electron mean free path (which is a function of the electron temperature T_e) are known, the drift velocity can be directly calculated and compared with that derived from the drift current equation. The comparison thus drawn should enable the physical properties of the spark channels to be further elucidated. Before the above calculation of v can be made, the following points should be raised.

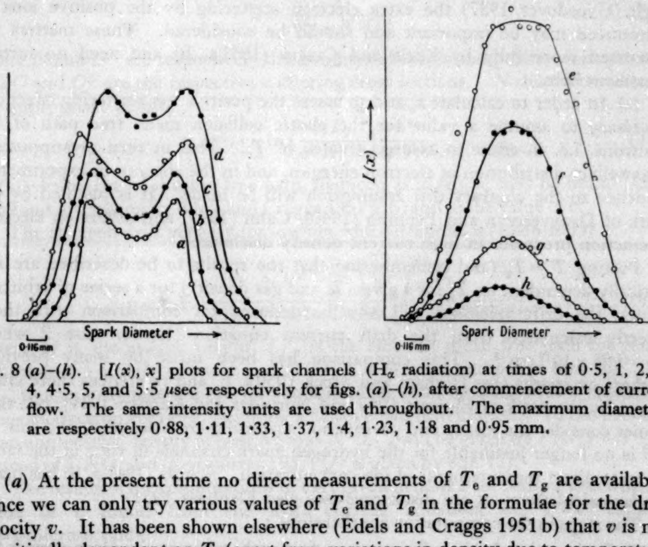


Fig. 8 (*a*)–(*h*). $[I(x), x]$ plots for spark channels (H_α radiation) at times of 0·5, 1, 2, 3, 4, 4·5, 5, and 5·5 μsec respectively for figs. (*a*)–(*h*), after commencement of current flow. The same intensity units are used throughout. The maximum diameters are respectively 0·88, 1·11, 1·33, 1·37, 1·4, 1·23, 1·18 and 0·95 mm.

(*a*) At the present time no direct measurements of T_e and T_g are available, hence we can only try various values of T_e and T_g in the formulae for the drift velocity v. It has been shown elsewhere (Edels and Craggs 1951 b) that v is not so critically dependent on T_g (apart from variations in density due to temperature changes in a constant pressure régime) as on T_e and so, if we assume $T_e = T_g$ the resulting errors will not be large (see Edel and Craggs 1951 b, table 2).

It may not be immediately apparent that a close approach to equilibrium between the electron and atom swarms ($T_e \simeq T_g$) in a discharge lasting only a few microseconds may be obtained. The relaxation time for the process is discussed by Margenau *et al.* (1946) and Weizel and Rompe (1947) and the former authors point out that since an electron loses on an average $2\,\mathbf{m}/M$ of its energy per elastic collision it must make $M/2\mathbf{m}$ collisions to lose an appreciable part of its energy. The time between collision is λ/u where u is the random velocity of an electron and so the relaxation time is roughly

$$\frac{M}{2m}\frac{\lambda}{u} \sim \frac{920 \times 10^{-4}}{2 \times 10^{6}} \sim 10^{-7} \text{ sec.}$$

λ here includes the effect of positive ion scattering (Gvosdover 1937) discussed below. This assumes that $\Delta T_e \gg \Delta T_g$, i.e. that T_e falls rapidly from a value very much higher than T_g to a much lower value which is still so great that T_g is negligible compared with T_e. This probably applies to the earliest parts of a spark discharge. The fractional energy loss for an electron per elastic collision is, more accurately, $(8m/3M)(1-T_g/T_e)$. If, however, the gas is heated by elastic collisions with an electron swarm whose temperature never greatly exceeds the value of T_e at, say, 3 μsec then the appropriate relaxation time is $(N_0/N_e)(M/2m)\lambda/u$ (where N_0 and N_e are gas atom and electron particle densities) or, since $N_0 \sim 10 N_e$ for the present case, of the order of 10^{-6} second.

(*b*) The electron mobility (k_e) formula (Loeb 1939) must be correctly used for the present case when $T_e \sim T_g$, and since the ion concentration is sufficiently high (Gvosdover 1937) the extra electron scattering by the positive ions so introduced may be important and should be considered. These matters are discussed more fully by Edels and Craggs (1951 a, b) and need no further treatment here.

(*c*) In order to calculate k_e and to assess the positive ion scattering effect it is necessary to assume a value for the elastic collision mean free path of the electrons, i.e. in effect to assume a value of T_e. This in turn presupposes a Maxwellian distribution of electron energies, and in the absence of experimental evidence to the contrary this assumption will be made. It is justified by the work of Druyvesteyn and Penning (1940), Cahn (1949) and others on electron interaction processes in high current density discharges.

Putting $T_e = T_g$ (and remembering that the results to be described are not critically dependent on T_g for a given T_e and gas density) for a series of arbitrary values, the drift velocities will now be deduced for comparison with those directly calculated from the drift current equation (1), for case 2 where $N_e = 0{\cdot}95 \times 10^{17}$ cm^{-3}. This comparison has been made by many previous workers on sparks (see Higham and Meek 1950 a, b, and references there cited, especially the work of Flowers 1943 and Norinder and Karsten 1949), but they do not consider the elastic and positive ion scattering cross sections, and this we feel is no longer justifiable for the hydrogen spark channels in view of the fairly accurate direct measurements of ion concentration. Schulz (1947 a, b) has made similar calculations on arc discharges (see also Elenbaas 1951).

The steps in the computation are therefore :

(i) Deduce the elastic collision mean free path, taken at the mean energy for the given value of T_e and using the theoretical Ramsauer cross sections for atomic hydrogen (communicated privately to us by Professor Massey). The value of T_g ($=T_e$) gives the particle density necessary for the above calculation, if the spark channel is considered at $t = 2\,\mu$sec (Case 2) to be at atmospheric pressure. If the constant density régime is considered to persist to this time*, then the particle concentration is twice atmospheric value because of virtually complete dissociation. The shock wave work will be discussed elsewhere. Calculations are given in table 2 for both régimes, which are, at least, the possible extremes.

---

* This has been usually the case in previous spark work, e.g. Craggs and Meek (1946), Norinder and Karsten (1949), Higham and Meek (1950 a, b) but was first considered doubtful by one of us (R.D.C.) after consideration of the initial expansion process as a shock wave mechanism.

The elastic cross sections for hydrogen, used in the calculations of table 2 are given in table 1.

Table 1

Electron energy (ev)	0·135	0·30	0·54	1·2	2·2	3·4	4·9	8·6	13·5	19·5	30·4
P_c	$1{\cdot}76\times10^2$	$1{\cdot}42\times10^2$	$1{\cdot}13\times10^2$	76	54	40	30	17·6	10·5	6·5	3·5

P_c=effective area of all atoms in 1 cm³ at unit pressure. If $p=1$ mm Hg (0°c) then the effective area/atom is q where $q=0{\cdot}28\times10^{-16}P_c$ cm².

(ii) Correct the above mean free path in accordance with the Gvosdover theory, i.e. for positive ion scattering (Schulz 1947 a, b, Elenbaas 1951, Druyvesteyn and Penning 1940). The corrected mean free path is then given by

$$\lambda=\frac{1}{NQ+N^+Q^+} \qquad \ldots\ldots(2)$$

where N and N^+ are respectively the numbers of atoms and positive ions per cm³ and Q and Q^+ are the respective scattering cross sections. N should be taken at the appropriate density.

$$Q^+=\frac{\pi}{2}\ln\left(\frac{\boldsymbol{k}T_e}{\boldsymbol{e}^2N^{+1/3}}\right)\frac{\boldsymbol{e}^4}{\boldsymbol{k}^2T_e^{\,2}}. \qquad \ldots\ldots(3)$$

(iii) From the final mean free path, deduced as in (ii), k_e may be found (Loeb 1939 and discussion by Edels and Craggs 1951 a, b) and so the value of v.

The electron drift velocity for low ion concentrations is (Loeb 1939)

$$v=k_eX=\frac{271\times10^3X\left(\dfrac{273}{T_g}\right)^{1/2}L_{1,0}}{[1+\{1+1{\cdot}1\times10^6ML_{1,0}{}^2(X/p)^2\}^{1/2}]^{1/2}} \qquad \ldots\ldots(4)$$

where M is the atomic weight, $L_{1,0}$ is the electron mean free path for elastic collisions at 1 mm Hg pressure and 273°K, X is in v/cm, and p is in mm Hg. For $L_{1,0}$ we can only take a mean for the Ramsauer cross sections over a reasonable range of electron energies, assuming a Maxwellian distribution. In the presence of an appreciable density of positive ions N^+ $L_{1,0}$ will be given by $1/(NQ+N^+Q^+)$ instead of $1/NQ$, and (Gvosdover 1937) Q^+ is given by eqn. (3).

The results of the computations of v are given in table 2 for comparison with the value of $\boldsymbol{v}$ found from the drift current equation, i.e. 7×10^5 cm sec⁻¹ ($\boldsymbol{t}=2$ microsecond).

Table 2

$T_e=T_g$ (°K)	NQ (cm²/cm³)	N^+Q^+ (cm²/cm³)	v (cm sec⁻¹)
(i) Constant (atmospheric) pressure channel			
10000	$1{\cdot}5\times10^3$	11×10^3	$2{\cdot}6\times10^5$
15000	$0{\cdot}8\times10^3$	$5{\cdot}6\times10^3$	$0{\cdot}8\times10^5$
20000	$0{\cdot}48\times10^3$	$3{\cdot}4\times10^3$	$1{\cdot}4\times10^6$
(ii) Constant density channel			
10000	11×10^4	11×10^3	$0{\cdot}9\times10^3$
15000	$8{\cdot}6\times10^4$	$5{\cdot}6\times10^3$	$1{\cdot}8\times10^3$
20000	7×10^4	$3{\cdot}4\times10^3$	$2{\cdot}1\times10^3$

It will be seen that, for the constant pressure régime, a channel temperature of almost 15 000°K will give v in agreement with the drift current value, but that the constant density case results in very low values of v and would necessitate unreasonable values of T_g to give v of the order of 10^5 cm sec^{-1}.

The above treatment refers to the case of atomic hydrogen, since spectroscopic work has failed to reveal any appreciable molecular hydrogen emission. Figure 10 (Plate II) gives a spectrum of the discharge which shows only the radiation in the Balmer series of atomic hydrogen, greatly widened by the Stark effect which we have used (Craggs and Hopwood 1947 a) to give the estimates of electron density used in the drift current equation above. Figure 10 shows no evidence of the spectrum molecular hydrogen and the H_α line is greatly over-exposed. With even longer exposures evidence of weak spectral continua, partly due perhaps to bremsstrahlung, are obtained. Measurements of the relative intensity of the H_α, H_β and H_γ lines, as in fig. 10, have been taken with photomultiplier tubes, in order to investigate the excitation mechanisms in hydrogen spark channels and will be discussed elsewhere.

Clearly, tungsten vapour from the electrodes would affect the results if allowed to diffuse into the region of the channel being observed, but as the initial velocity of propagation of the electrode vapour clouds into the channel is of the order of 10^5 cm sec^{-1}, and falls below this figure further from the electrodes, the centre of a 1 cm gap will remain virtually free of vapour for 5 μsec or more (Williams *et al.* 1949).

The spark temperature, on the assumption of thermal equilibrium, can be deduced from Saha's equation. If the ionization potential $V_1 = 13{\cdot}5$ ev, and the mean $N_e = 0{\cdot}95 \times 10^{17}$ cm^{-3} then the Saha temperatures for the cases of constant pressure and constant density are respectively 13 500 and 9400°K. In the latter case the particle density is doubled to allow for complete dissociation and, in this case, the Saha temperature is inconsistent with the temperature deduced from table 2.

Kirschstein and Koppelmann (1937) have, in a similar manner, estimated the temperature of certain arc discharges, in high pressure gas streams, from current and electron densities and found values of about 15 000° with current densities of the order of 20 000 to 50 000 A cm^{-2}, to be compared with our current densities of about 9000 A cm^{-2}. References to other earlier work in this field have been given by Craggs and Hopwood (1947 a) and Edels (1950).

Further indirect information on the channel gas temperature is available if the central fall in channel brightness shown in the records of fig. 8 (*a*) for example, is indeed due to the gas being so near to complete ionization that the concentration of excited atoms falls; this needs practically 100% ionization since the degree of excitation at these temperatures is only of the order of 10^{-4}. It is by no means certain that this explanation is correct but if the constant pressure régime is reached quickly, the particle density will be low enough ($T_g = 10\,000$–$20\,000$°K) for N_e of the order of 10^{17} cm^{-3} to be equivalent almost to 100% ionization. This simple argument seems to suggest, then, either that the constant pressure state is reached in two microseconds or less, or that some other mechanism explains the low central channel intensity.

The central low density trough is a consequence of the initial shock wave expansion mechanism and the relevant theory has been discussed by various authors, notably G. I. Taylor (1950).

The measurements of channel voltage gradient (figs. 5 and 7) should, it was hoped, enable the mean electron energy to be calculated and this information would be of great value in considering recombination processes. Davydov (1935) has given a treatment by which the mean energy of the electrons in a discharge swarm can be calculated when $T_g \sim T_e$. Usually (see Druyvesteyn and Penning 1940 and Llewellyn Jones 1944 for references) the effect of T_g need not be, and is not, considered in low density, low current discharges. Davydov's method uses a constant electron mean free path, which is clearly only an approximation, and gives the mean squared electron velocity as

$$\overline{v^2} = \frac{\int_0^\infty f_0 v^4 \, dv}{\int_0^\infty f_0 v^2 \, dv} \qquad \ldots\ldots(5)$$

where $f_0 = (A+B)^B \exp -A$, with $A = \boldsymbol{m} v^2/2\boldsymbol{k}T_g$ and $B = \frac{1}{6}(M/\boldsymbol{m})(\boldsymbol{e}Xl/\boldsymbol{k}T_g)^{1/2}$; M and $\boldsymbol{m}$ are the atom and electron masses, and l the electron mean free path. The remaining symbols have their usual significance. It will be noted that l (and so T_e) and T_g must be known and this information is not available. We shall put $T_g = 15\,000°\text{K} = T_e$ and deduce l from T_e (Maxwell distribution) correcting for the positive ion scattering as above, and taking $X = 90$ v/cm. Tedious graphical integration gives the mean electron energy as 2·1 ev, i.e. practically the same as would be derived from $T_e = T_g$ for $X \to 0$. Note that this result is not critically dependent on l, i.e. on T_e, but more so on T_g. Indeed the exponential term in f_0 above which does not involve X is of much greater importance than the res tof f_0, which for our case is approximately equal to unity. Davydov's formula in fact gives a Maxwell distribution if $\boldsymbol{e}Xl(M/\boldsymbol{m})^{1/2} \gg \boldsymbol{k}T_g$, and a Druyvesteyn distribution if $\boldsymbol{k}T_g \ll \boldsymbol{e}Xl(M/\boldsymbol{m})^{1/2}$, but in these spark channels $\boldsymbol{e}Xl(M/\boldsymbol{m})^{1/2} \sim \boldsymbol{k}T_g$.

The channel expansion curves of fig. 9, which relate to the visibly luminous discharge, may be further discussed. We shall treat the initial (rapid) expansion process elsewhere but on the assumption that the channel in its later stages expands by diffusion its radial growth may be roughly calculated. We assume that the ambipolar diffusion of the charged particles determines the channel size, i.e. that the luminosity is accompanied by free electrons, as it would be if excitation by electrons or radiative electron/ion recombination were responsible for the light emission.

Then the channel diameter d will change with time t through the relation (Wilson 1951)

$$2r = d = 4{\cdot}7(D_a t)^{1/2} \text{ (for the diameter to 90\% fall-off of intensity)}, \qquad \ldots\ldots(6)$$

where D_a is the ambipolar diffusion coefficient. The calculation of D_a for atomic hydrogen at the high (unknown) channel temperature is difficult but the value for helium, and its temperature variation (Biondi and Brown 1949) may be taken as an approximation. At 15 000°K, D_a is then approximately 2×10^3 cm² sec⁻¹ and dr/dt for the channel approximately 2×10^{-2} cm/μsec, to be compared with the experimental value (fig. 11) of approximately 5×10^{-2} cm/μsec. Figure 11 shows that r varies as $t^{1/2}$ after $t \simeq 3\,\mu$sec, as eqn. (6) demands. Also, for the initial part of the expansion, say for the times 0–2 μsec in fig. 9, the theory of Taylor (1950) and Drabkina (1951) is applicable. This is when the expansion of the luminous channel coincides with the shock wave generated during the early part of the spark channel formation process. The rate of expansion however is then much greater than it is during the later, diffusion, stage, i.e. the constants relating r^2 to t in the two régimes are different. The resolution of the rotating mirror camera used in the present work is hardly adequate for satisfactory resolution of the first 0·5 μsec, as shown by the blurred envelope at the beginning of the streak photographs of fig. 1. It must however, be emphasized that the importance of thermal excitation during and after current flow has still to be assessed and the relative contributions of electron/atom and atom/atom collisions (von Engel and Steenbeck 1934) to any such thermal excitation will need to be determined. If thermal excitation is important, with atom/atom collisions predominating, then the operative diffusion (see Edels and Craggs 1951 a, b) process will be that of the 'hot' atoms in the channel.

The fact (fig. 9) that the visible channel contracts at the end of the current pulse suggests, if recombination is responsible for the radiation, a lower limit for the recombination coefficient, assuming D_a to be known as above. Thus, on these assumptions, for the skirts of the channel to cease radiating $\alpha N^2 > D_a(\partial N/\partial r)$ where α is the recombination coefficient and N the electron concentration at the periphery of the channel. Taking N and $\partial N/\partial r$ from unpublished work of Hopwood *et al.*, $\alpha \sim 10^{-10}$. Further measurements will be made to investigate the validity of the above assumptions.

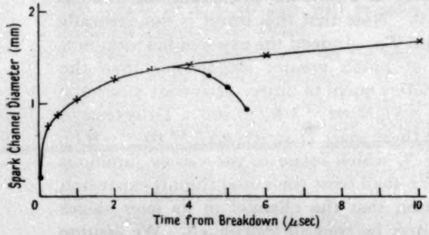


Fig. 9. The variation of spark channel diameter with time for 4 and 10 μsec current pulses.

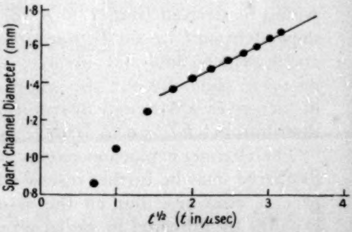


Fig. 11. The variation of spark channel diameter with time.

§ 5. CONCLUSIONS

Measurements of the voltage gradient and channel cross section for hydrogen spark channels, carrying about 100 amperes for several microseconds, have been made. The results are consistent with a channel gas (or electron) temperature (T_g or T_e) of about 15 000°K and with the establishment, in times less than 2 microseconds, of constant pressure (arc-like) conditions, in the channels. A knowledge of T_g and T_e from direct experiments would be valuable and such work is being undertaken.

ACKNOWLEDGMENTS

We wish to thank many colleagues, especially Professor J. M. Meek, Mr. W. Hopwood, Dr. H. Edels, Dr. J. B. Higham and Mr. Tsui Fang for discussions. One of us (R.D.C.) is indebted to the University of Liverpool for a Kitchener Studentship. We are also grateful to Professor H. S. W. Massey for reading the paper.

REFERENCES

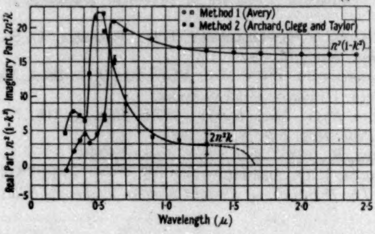
BIONDI, A. M., and BROWN, S. C., 1949, *Phys. Rev.*, **75**, 1700.
CAHN, J. H., 1949, *Phys. Rev.*, **75**, 293.
CRAGGS, J. D., and HOPWOOD, W., 1947 a, *Proc. Phys. Soc.*, **59**, 755; 1947 b, *Ibid.*, **59**, 771.
CRAGGS, J. D., and MEEK, J. M., 1946, *Proc. Roy. Soc.* A, **186**, 241.
CRAGGS, J. D., WILLIAMS, G. C., and HOPWOOD, W., 1948, *Phil. Mag.*, **39**, 329.
DAVYDOV, B., 1935, *Phys. Z. Sowjet*, **8**, 59.
DRABKINA, S. I., 1951, *J. Exp. Theor. Phys. U.S.S.R.*, **21**, 473.
DRUYVESTEYN, M. J., and PENNING, F. M., 1940, *Rev. Mod. Phys.*, **12**, 87.
EDELS, H., 1950, E.R.A. Report No. L/T 230.
EDELS, H., and CRAGGS, J. D., 1951 a, *Proc. Phys. Soc.* A, **64**, 562; 1951 b, *Ibid.*, **64**, 574.
ELENBAAS, W., 1951, *The High Pressure Mercury Vapour Discharge* (Amsterdam: North Holland Publishing Co.).
VON ENGEL, A., and STEENBECK, M., 1934, *Elektrische Gasentladungen* (Berlin: Springer).
FLOWERS, J. W., 1943, *Phys. Rev.*, **64**, 225.
GEGECHKORI, N. M., 1951, *J. Exp. Theor. Phys. U.S.S.R.*, **21**, 493.
GVOSDOVER, S. C., 1937, *Phys. Z. Sowjet*, **12**, 164.
HIGHAM, J. B., and MEEK, J. M., 1950 a, *Proc. Phys. Soc.* B, **63**, 633; 1950 b, *Ibid.*, **63**, 649.
HOLTHAM, A. E. J., and PRIME, H. A., 1950, *Proc. Phys. Soc.* B, **63**, 561.
JONES, F. LLEWELLYN, 1944, *Proc. Phys. Soc.*, **56**, 239.
KIRSCHSTEIN, B., and KOPPELMANN, F., 1937, *Wiss. Veröff. Siemens-Werk.*, **16**, 1, 51; **16**, 3, 26.
LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York: Wiley).
MARGENAU, H., McMILLAN, F. L., DEARNLEY, I. H., PEARSALL, C. S., and MONTGOMERY, C. G., 1946, *Phys. Rev.*, **70**, 349.
NORINDER, H., and KARSTEN, O., 1949, *Ark. Mat. Astr. Fys.*, **36** A, Part 4.
SAXE, R. F., 1948, *Thesis*, University of Liverpool.
SCHULZ, P., 1947 a, *Ann. Phys., Lpz.*, **1**, 95; 1947 b, *Ibid.*, **1**, 318.
TAYLOR, G. I., 1950, *Proc. Roy. Soc.* A, **201**, 159.
WEIZEL, W., and ROMPE, R., 1947, *Ann. Phys., Lpz.*, **1**, 285.
WILLIAMS, G. C., CRAGGS, J. D., and HOPWOOD, W., 1949, *Proc. Phys. Soc.* B, **62**, 49.
WILSON, J. G., 1951, *The Principles of Cloud Chamber Technique* (Cambridge: University Press).

LETTERS TO THE EDITOR

The Optical Constants of a Single Crystal of Germanium

There has recently become available to us through the kindness of Mr. I. Cressell of the Marconi Wireless Telegraph Co., Ltd., Chelmsford, what appears to be a unique single crystal of germanium. This was grown by cooling a molten ingot and has a hexagonal [111] face nearly a perfect plane (resistivity 8 ohm cm). Because of the size of this face and the perfection of the lattice as shown by x-ray examination it seemed desirable to measure the optical properties of the crystal.

Two different reflection methods were used for the measurements in two wavelength regions. For wavelengths from $2{\cdot}5\,\mu$ to $0{\cdot}44\,\mu$ the method recently described by one of us (Avery 1952) was used, and in the ultra-violet from $0{\cdot}44$ to $0{\cdot}25\,\mu$ measurements were made by one of the techniques described by Archard, Clegg and Taylor (1952). Facilities to carry out the latter were made available to the authors by the generous co-operation of Professor A. M. Taylor of the University of Southampton.



Graphs of $n^2(1-k^2)$ and $2n^2k$ against wavelength.

The results obtained are shown in the accompanying graph, in which the real and imaginary parts of the dielectric constant are plotted against wavelength. The variation of the imaginary part shows that in addition to the strong absorption band at $0{\cdot}52\,\mu$, already reported by Brattain and Briggs (1949), a second and less intense band exists with its peak at a wavelength of $0{\cdot}31\,\mu$. The existence of an absorption band in this region has already been reported in evaporated layers of germanium by Gebbie (1952), although he found the strength of the band very much greater than that noted here.

The refractive index of germanium in the wavelength range $2{\cdot}0$–$2{\cdot}5\,\mu$ as determined from these measurements is $4{\cdot}00 \pm 0{\cdot}05$. This compares with the values of $4{\cdot}10$ at $2\,\mu$, and $4{\cdot}07$ at $2{\cdot}6\,\mu$, obtained by Briggs (1950) from measurements using germanium prisms.

One interesting feature of these results concerns the long wavelength tail to the main absorption band. For a number of reasons the techniques used for the measurements in this region are not capable of a high accuracy in the determination of small values of the imaginary part. Whilst the values shown for $2n^2k$ at $1{\cdot}1\,\mu$ and $1{\cdot}3\,\mu$ may be in error by as much as a factor of two, it appears that there remains considerable absorption in the $1{\cdot}3\,\mu$ region, and that this absorption is not decreasing rapidly with wavelength. The transmission experiments of Fan and Becker (1950) show a sharp fall in the absorption coefficient of pure crystalline germanium in the region $1{\cdot}6$–$1{\cdot}7\,\mu$, and if this observation is coupled with the present data, the structure of the long wavelength tail of the band may be as shown dotted in the figure. The similarity with the form found for the long wavelength tail of the compounds PbS, PbSe and PbTe (Avery 1953) is marked, and may suggest some similarity in the band structures.

Calculations using classical dispersion theory show that, in contrast to the data for the lead compounds, these data are not self-consistent, i.e. if the theory holds there must exist further absorption bands at even shorter wavelengths to account for the high value of the dielectric constant in the infra-red. However, it is by no means certain that the simple theory need apply. From the shape of the absorption band at $0{\cdot}52\,\mu$ a value of $0{\cdot}8$–$1{\cdot}0$ can be given for the oscillator strength.

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this note.

Telecommunications Research Establishment, D. G. AVERY.
Great Malvern, Worcs.
The University, Southampton. P. L. CLEGG.*
3rd February 1953.

* Now at Imperial Chemical Industries Ltd. (Plastics Division), Welwyn Garden City, Herts.

ARCHARD, J. F., CLEGG, P. L., and TAYLOR, A. M., 1952, *Proc. Phys. Soc.* B, **65,** 758.
AVERY, D. G., 1952, *Proc. Phys. Soc.* B, **65,** 425; 1953, *Ibid.*, **66,** 134.
BRATTAIN, W. H., and BRIGGS, H. B., 1949, *Phys. Rev.*, **75,** 1705.
BRIGGS, H. B., 1950, *Phys. Rev.*, **77,** 287.
FAN, H. Y., and BECKER, M., 1950, *Semi-Conducting Materials*, ed. H. K. Henisch (London : Butterworths Scientific Publications).
GEBBIE, H. A., 1952, *Thesis*, University of Reading.

The Role of Convergence in Stereoscopic Vision

In an article that appeared in the *Proceedings of the Physical Society* (Wright 1951), data were presented which were interpreted as evidence that convergence movements of the eyes (apart from retinal disparity) are an important factor in stereoscopic depth perception. From this writer's point of view the data do not warrant this conclusion. In the experiment reported, a subject views two laterally separated apertures (points) in the dark at a distance of 6 feet. The distance of one of the two points from the subject can be varied. The threshold data for the perception of the difference in distance of the two points were determined : (*a*) when the subject looked back and forth from one of the points to the other, and (*b*) when one point was constantly fixated. The thresholds for difference in depth were found to be lower when eye movements (and therefore convergence movements) were permitted than when they were not permitted. Now in case (*b*), in which one point was constantly fixated, the stereoscopic perception would depend upon the stimulus of a peripheral retinal disparity only, and the threshold of stereo-acuity would be related to the visual acuity for that particular peripheral part of the two retinas. In case (*a*), in which the eyes look successively from one point to the other, there will be an optimal stereo-acuity when the point of convergence is halfway between the two object points. This follows because at that convergence the retinal images of the two points will fall on retinal areas that give an *average* maximum visual acuity. Hence the optimal stereo-acuity will depend upon the visual acuity at peripheral angles corresponding to only *half* the angular separation of the two points. Now if one plots the threshold data (slopes of the probability function) for case (*a*) upon the same graph as those for case (*b*), but at only half the angular separation of the object points, the two sets of data will fall essentially on the same curve, certainly within experimental error. This result shows then on the contrary that there is no influence of convergence movements as such on stereoscopic depth perception, that stereopsis is fundamentally only a function of disparity.

Langlands' failure to find stereoscopic perception in extramacular vision in flash illuminations is not evidence that disparity alone cannot be the stimulus for stereopsis. The results of Hertel and Monjé (1947), and of Vilmar (1950), show that there is a time–depth threshold for stereoscopic vision, and this threshold increases toward the peripheral retina. The flash illumination and the depth interval used by Langlands probably were below these thresholds.

Finally it should be pointed out that the fact that stereopsis can occur from double images was well known to Helmholtz and to Hering. The latter's ' dropping bead test ' gave convincing proof of this. However, there are limits of disparity within which stereopsis can occur with double images. Measurements of these limiting disparities have recently been made (Ogle 1952).

Mayo Clinic,
Rochester, Minnesota, U.S.A.
4th March 1953.

K. N. Ogle.

Hertel, K., and Monjé, M., 1947, *Arch. ges. Physiol.*, **249**, 295.
Ogle, K. N., 1952, *J. Exper. Psychol.*, **44**, 253.
Vilmar, K. F., 1950, *Klin. Monatsbl. f. Augenh.*, **117**, 242.
Wright, W. D., 1951, *Proc. Phys. Soc.* B, **64**, 289.

Dr. Ogle is a leading authority on binocular vision and his comments on my paper are very welcome, especially as my own excursions into the subject as an active field of research are of comparatively recent origin. His alternative interpretation of my experiments, however, rests on the assumption either that the brain utilized the visual information obtained during the rapid sweep of the eyes as they transferred their fixation from one target to the other, which seems extremely unlikely, or else that a fixation pause occurred midway between the two targets.

The effect of such a pause was not overlooked, as the following quotation from the original paper will show : " The observer was asked to try to transfer his gaze from P to Q in one fixation jump, and certainly without any fixation pauses midway between P and Q. . . . The observer was definitely discouraged from looking backwards and forwards several times, otherwise the risk of a fixation pause between P and Q would have been too great ".

These instructions may not always have prevented fixation pauses, and an objective record of the eye movements during the experiment to check this was hardly feasible. Nevertheless, on most occasions, it can reasonably be assumed that fixations near the mid-point did not occur, an assumption which is supported by the ' blind spot ' experiment. For if the acuity for condition (*a*) were in fact due to observations made when the convergence point was midway between the two targets, then the blind spot would not have affected the acuity at $\theta = 14°$, as in fact it did. Moreover, the fatigue effect which only intervened in the blind spot condition also confirms that the depth perception was not an artefact arising from fixation pauses intermediate between P and Q.

For these reasons, it is impossible to accept Dr. Ogle's conclusion that convergence has no influence on stereoscopic depth perception. With regard to his comment on stereopsis in the presence of diplopia, no claim was made that this was an original discovery; indeed, references were given to earlier work. But the surprise expressed when this effect was pointed out, has been so great that the phenomenon has obviously been overlooked by many people. These confirmatory experiments with different apparatus and under different observing conditions from those used in the past have certainly been of value to me in emphasizing the reality of what is after all a rather surprising result.

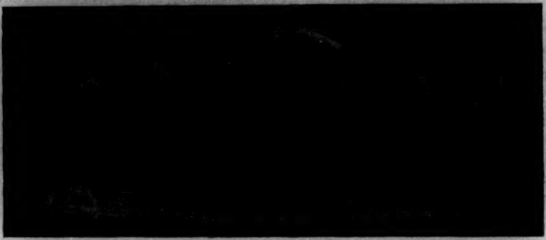
Technical Optics Section,
Imperial College, London.
23rd March 1953.

W. D. Wright.

Room Temperature Brittleness of Chromium

The room temperature brittleness of chromium has become a problem of considerable importance in view of the current interest in chromium-base alloys for high temperature applications. Previous investigators (Kroll, Hergert and Yerkes 1950, Greiner 1950, and Gilbert, Johansen and Nelson 1952) have successfully hot-worked pure chromium, but the material produced was almost invariably brittle below temperatures ranging from 100–300°c. No convincing explanation of this behaviour has so far been advanced.

In the present investigation chromium strip, varying in thickness from 0·010 in. to 0·030 in., was produced by forging, swaging and rolling arc-melted chromium ingots at a nominal temperature of 800–900°c. The chromium used was prepared by the electrolysis of chromic acid (Greenaway 1951) and had the following chemical analysis: oxygen, 0·06% by weight; nitrogen, $<0{\cdot}001\%$ by weight; carbon, $<0{\cdot}005\%$ by weight. No metallic impurities could be detected by spectrographic analysis. Contamination of the chromium from the atmosphere during working was avoided by sealing the billets in mild steel and/or stainless steel sheaths at all stages. Strip produced in this manner was tested in a simple bend test and some samples showed considerable ductility at room temperature whilst others appeared to be completely brittle. Microscopic and chemical tests revealed that contamination of the surface of the chromium had occurred from the mild steel envelope during rolling and, when this contaminated layer was removed by chemical dissolution, the strip was almost invariably ductile at room temperature. The contamination caused the surface layers to become hard and apparently brittle and this produced brittle fracture of the strip on bending. A further point was that the room temperature ductility of chromium does not appear to be related to grain structure, or certainly not to the extent that is found in the case of tungsten and molybdenum. Rolling at a nominal temperature of 900°C produced elongation of the grains in the rolling direction, but specimens with an equi-axed grain structure formed by annealing after the final pass through the rolls, were just as ductile at room temperature after removal of the contaminated surface layer, although in this instance the specimens had sometimes to be reduced in thickness to a greater extent before ductility was achieved. The room temperature ductility of these specimens was quite spectacular, and some samples could be bent through 180° and back again quite rapidly before fracture occurred. Some examples of chromium strip bent at room temperature are shown in the figure.



Some examples of chromium strip bent at room temperature.

Two conclusions can be drawn from this work: (*a*) chromium is not inherently brittle at room temperature, and (*b*) the brittleness of chromium appears to be associated with impurities in the metal. It would appear that the impurity causing the brittleness in the present experiments must have been absorbed from the mild steel envelope or from the air entrapped between the envelope and the specimen, so that the impurities most likely to be responsible are iron, carbon or nitrogen. Oxygen can almost certainly be discounted since there was a significant oxygen content in the original material and oxide inclusions were prevalent in the micro-structure of the rolled strips. Further investigation is proceeding to determine the specific impurity responsible. It is possible that the brittleness of chromium arises from the action of the impurity atoms in forming 'atmospheres' around dislocations (Cottrell 1948) and that chromium behaves at room temperature in a similar fashion to iron containing carbon and nitrogen at liquid air temperatures. Churchman and Cottrell (1951) have shown that iron single crystals, in which the dislocations were 'anchored' by carbon and/or nitrogen atoms, failed at the temperature of liquid air by twinning after a small amount of deformation. However, if the dislocations were not 'anchored', the single crystals deformed considerably by slip before fracture. This behaviour is attributed to the large increase in the temperature dependence of mechanical properties which is found in a yield point system as compared with a system in which no yield point is shown.

Thanks are due to the Chief Scientist, Department of Supply, Melbourne, Australia, for permission to publish this note, to Mr. H. T. Greenaway of these Laboratories for supplies of chromium, and to Professor J. Neill Greenwood, Research Professor of Metallurgy, University of Melbourne, for the use of a swaging machine.

Aeronautical Research Laboratories, H. L. WAIN.
Department of Supply, Melbourne, Australia. F. HENDERSON.
17th March 1953.

CHURCHMAN, A. T., and COTTRELL, A. H., 1951, *Nature, Lond.*, **167,** 943.
COTTRELL, A. H., 1948, *Report of Bristol Conference on Strength of Solids* (London : Physical Society), p. 30.
GILBERT, H. L., JOHANSEN, H. A., and NELSON, R. G., 1952, U.S. Bureau of Mines Report of Investigation 4905.
GREENAWAY, H. T., 1951, Department of Supply, A.R.L. Rept. S.M. 163.
GREINER, E. S., 1950, *Trans. Amer. Inst. Min. Metall. Engrs.*, **188,** 891.
KROLL, W. J., HERGERT, W. F., and YERKES, L. A., 1950, *J. Electrochem. Soc.*, **97,** 258.

REVIEWS OF BOOKS

Fundamentals of Engineering Electronics, by W. G. Dow. Pp. xviii + 627. 2nd edn. (New York : John Wiley; London : Chapman and Hall, 1952.) 68*s*.

It is becoming more and more difficult to define the scope of *electronics*. The subject today is concerned with the motions of electrons and ions under electric fields *in vacuo*, in gases, and also in metals, as well as with the important surface or boundary phenomena occurring at the electrodes. The fundamentals of electronics therefore involves a great deal of modern extra-nuclear statistical physics. The vast borderland between physics and light electrical engineering is of equal interest to the physicist and electrical engineer, and some knowledge of the fundamental physics of the phenomena of electronics is of very great importance to the modern electrical engineer. *The Fundamentals of Engineering Electronics* by William G. Dow was written primarily with the view of providing students of electrical engineering with a reasonably complete and satisfactory account of the fundamental physical processes required for the understanding of the operation of electronic devices, and it appears to the reviewer that this object is achieved. This volume is a revised and modified edition of the author's book published in 1937, and it certainly provides the electrical engineering student with an excellent introduction to basic physical principles and the more elementary quantum and statistical aspects of electronics.

The first three chapters are concerned with the mathematical treatment of the motion of electrons *in vacuo* under electric and magnetic fields, with applications of the analysis to ultra-high frequency systems, magnetrons and cathode-ray tubes. The next four chapters are devoted to a treatment of electrostatic fields and space charge currents by means of conformal transformation. Considerable use is made of potential energy diagrams throughout.

The short seventh chapter is devoted to thermionic cathodes and thermionic emission; no mention is made of the decay characteristics of pulsed thermionic emission.

The quantum and statistical concepts in the modern theory of metals are introduced in a clear and simple manner in Chapter VIII, and extended to the case of semiconductors. This second edition contains an excellent account of the energy level behaviour of semiconductors. After this the author returns to deal with thermionic emission in the presence of high fields (Schottky effect) and field emission of electrons is dealt with briefly : a more systematic treatment of the subject would perhaps have been desirable and the account of field emission is not really adequate in view of its great importance in discharge phenomena. The influence of these effects and of contact potentials in electronic valves is discussed. Mention should certainly be made of the excellent elementary account given of the ionization and excitation of atoms, and the electronic structure of atoms. This covers the ground necessary for introduction to the more specialist treatment of these subjects. Chapters IX and X deal with the design of amplifiers; as here a knowledge of fundamentals is not strictly necessary, there seems no very great justification for the inclusion of this matter in the book, since this aspect is adequately treated in many others which are devoted solely to the properties of electrical circuits. Chapters XII and XIII are of a more fundamental character and deal with Maxwell–Boltzmann statistics with applications to electron emission and noise.

Photoelectric devices and the amplification of currents by ionizing collisions between electrons and gas molecules are treated in Chapter XIV. It is perhaps unfortunate that the definition of Townsend's α given on p. 439 is incomplete : the words " in the direction of the applied field " should be added. Chapter XV treats qualitatively the development of the Townsend type discharge, glow and arc discharges, and is a good introduction to the subject.

In a volume which attempts such a comprehensive survey of the basic physical processes of the motions of electrons and ions in gases, it is unfortunate that it was not possible to bring up-to-date the comparatively short treatment of electrical breakdown on pages 510–513. The account given here is short and entirely qualitative, and, in the light of experimental work published since 1950, completely misleading to the student. The MKS system is used throughout. A highly pleasing feature of the book is the inclusion of a list of problems at the end of each chapter. A student who works through all these can certainly feel that he has understood the subject-matter. From the choice of question it would appear that the author is an accomplished teacher.

The book of 627 pages is well presented and bound, and apart from one or two printing errors (e.g. ' lies ' for ' lines ' on page 8) is up to the usual high standard of publication by Wiley. It will be found a most useful addition to the library of both physicists and electrical engineers as well as to production engineers, and it can be strongly recommended. F. LL. J.

The Essentials of Fluid Dynamics, by L. Prandtl. Pp. x + 452. (London : Blackie and Son, 1952.) 35*s*.

Lamb's monumental book has reigned supreme for a good long time in the subject of fluid dynamics. Since Lamb's first edition appeared in 1879, many other books have been written, mostly with the intention of providing accounts of the subject more suited to the needs of students, and nearly all of them have claimed to be up-to-date and to be in tune with current research. In fact, the weight of authority of Lamb's *Hydrodynamics* has laid heavily upon them, and the concessions to current interests have been of no great significance, consisting for the most part of the addition of one or two chapters at the end on topics like high speed flow, rather than a radical change in emphasis and outlook. It has been remarked that from reading Lamb's *Hydrodynamics* you would never realize that water was wet; and the text books written since 1879 likewise fail to convey the real facts about the motion of fluids. The point is that hitherto text books have given prior place in the presentation to the case of irrotational low-speed flow, doubtless because this case is mathematically tractable and presents analogies with other branches of mathematical physics. It is known that a good approximation to irrotational flow does occur in practice in certain circumstances, but it is also known that the theory of irrotational flow can never explain certain vital—and, as we now think, rather interesting—aspects of the flow of real fluids with viscosity and heat conductivity. The flow generated by a sphere moving through a real fluid (it is a revealing indication of the present position, that when referring to what happens in nature one feels obliged to say explicitly that the fluid is *real*) is radically different from that corresponding to the irrotational motion of a frictionless fluid, and analysis of the former case will always have more intrinsic interest for the physicist, however incomplete its mathematical development may be.

The book under review is an excellent account, probably the best of its kind now available, of the flow of real fluids (which may not have wetness, but do have the properties relevant to dynamics). Professor L. Prandtl is one of the most distinguished of living German scientists, and has long been noted for his gift of clear and penetrating exposition. His own contributions to fluid dynamics have been the starting point for innumerable researches (his paper putting forward the theory of the viscous boundary layer provides an obvious example), and many able men have studied in his laboratory at Göttingen. Prandtl's elementary account of fluid flow in *The Physics of Solids and Fluids* is well known and appreciated among students of the subject, and the present book has grown from this by a number of stages. In its present form the book under review is a translation of the third edition of a work first published in Germany in 1942. It has retained the style of the author's chapters in *The Physics of Solids and Fluids* and has a very much wider scope.

The book consists of four different and effectively self-contained chapters, preceded by a short introductory discussion of pressure relations applying to fluids in equilibrium. The first of these chapters is concerned with the dynamics of frictionless fluids, and by a careful choice of material the author has managed to put into 60 pages all the essential results, including a number (such as the very important momentum theorems for the fluid within a ' control surface ', used to find expressions for the drag or thrust forces exerted by bodies moving through fluid) which are usually neglected in text books. The second describes the motion of viscous fluids, and the associated phenomena of boundary layers and turbulence; the treatment is concise, very much to the point, and provides good reading for either the student or the specialist. Also included in this chapter is an account of the theory of aerofoils and of related devices, such as propellers, for producing thrust on a fluid. The third section deals with the high speed flow of gases, not in great detail and, as elsewhere, avoiding mathematics where possible, but nevertheless conveying the essential physical ideas involved in the flow of a compressible fluid. The last chapter is on miscellaneous topics which, as the author remarks, " do not fit conveniently into the previous chapters but cannot be omitted from a comprehensive account of hydrodynamics "; *cannot* is the right word, despite the fact that, to my knowledge, no other book on fluid dynamics provides a description of cavitation, flow of solid–fluid mixtures, bodies in accelerated flow, rotating fluids, and in particular the effect of the earth's rotation on currents in the atmosphere and oceans, flow of stratified fluids, with applications to meteorology, forced and free convection, and many other phenomena. It is in this last chapter that the breadth and penetration of Prandtl's understanding of his chosen subject really emerges.

Now that proper tribute has been paid to this unique book, perhaps one adverse comment may be permitted. There seems to be little mention of the extensive developments that have occurred since the date of the earlier edition in 1942 and very little, in particular, about English and American work. This is hardly a criticism, for it is only the first edition that one has a right to expect to be up-to-date, and I mention it chiefly to warn prospective readers. An appendix of 90 notes and references to papers published in Great Britain and U.S.A. since 1939, prepared by Professor H. B. Squire and others, is given at the back of the book, but although they are well selected it remains true that the book gives an incomplete picture of the present position of research.

To summarize, this is a book which can be read with profit by either students or specialists, either as a short reference work or as a book for browsing. Students of applied mathematics will find perhaps that there is insufficient formal theory for their needs, but all who find pleasure in neat explanations of physical phenomena, and who like to hear about new phenomena, will be well satisfied. The book fills admirably the long-felt need for a description of fluid flow as it is in nature. G. K. BATCHELOR.

A Selection of Tables for use in Calculations of Compressible Airflow, by L. ROSENHEAD and others; prepared on behalf of the Aeronautical Research Council by the Compressible Flow Tables Panel. Pp. viii + 143. (Oxford : Clarendon Press, 1952.) 40*s*.

Calculations of the various quantities describing the state of a gas which is moving at high speeds (that is, at speeds not small compared with the velocity of sound in the fluid) have become a routine operation for aeronautical and ballistic experts in recent years, and these tables have been prepared as an aid to such calculations. Some selections of tables of the required kind already exist, but the Panel appointed by the Aeronautical Research Council found them unsatisfactory and accordingly it set out to provide an " authoritative selection of the more important tables dealing with compressible airflow, collected under one cover and guaranteed, as far as possible, to be free from error ". All the indications are that it has succeeded, and the result is a well-printed volume of tables, easy to consult and to interpret. A companion volume of graphs of the tabulated quantities will appear soon, with a different size of page.

The tables are grouped under six headings; Isentropic flow tables, giving the values of quantities derivable from Bernoulli's equation when particles of the fluid move without change of entropy; Characteristic tables, used for the calculation of supersonic flow by the method of characteristics; Shock tables, giving the relations between the states of a gas on the two sides of a shock wave; Tables for the reduction of pressure ratios, which are needed for the interpretation of the readings of various pressure-measuring instruments; Tables of some algebraic functions occurring in calculations of high-speed flow; and Miscellaneous tables, mostly concerned with the physical properties of air. In all these tables the value of the ratio of the specific heats is taken as 1·4, in order to make them apply accurately to the choice of air as the fluid in motion. Readers who are not engineers may be irked a little by the use of the slug as the unit of mass. G. K. BATCHELOR.

Niederdruck-Stromrichterventile, by H. BERTELE. Pp. xii + 239. (Vienna: Springer-Verlag, 1952.) 66*s*. 6*d*.

The advances which have been made in static current-rectifying devices during the last few decades has rendered indispensable their use in many branches of electrical engineering. Their application, not only for the supply of direct current for the electrochemical and metallurgical industries, and for the speed control of motors, but also to other control systems is an indication of their versatility. The present book deals with some of the latest developments and is concerned mainly with low pressure current converters.

In the first chapter the author has confined the subject matter to some of the more important features of current–voltage relations and waveforms, commutation oscillations, and a brief account of the high pressure Marx type of rectifier. He has wisely omitted details of rectifier transformers and circuits which are treated extensively elsewhere. The second chapter is devoted to some of the fundamental physical processes involved in the low pressure spark discharges and includes sections on the elementary kinetic theory of gases, diffusion of ions and ionization and recombination processes. Basic physical processes which have important applications in rectifiers, viz. the Townsend theory of the low pressure spark discharge, plasma theory, wall effects and the Langmuir–Child equation are all considered.

The important role of the cathode in supplying the electrons is also treated in detail in this chapter and electron emission from cold, thermionically heated, and incandescent cathodes is considered. A very brief account is given of field emission and other possible mechanisms of electron emission from the cathode spot in pool type rectifiers. This chapter contains one small error: the equation II, 48 is due to Sommerfeld and Bethe ; it is a modification of the equation given earlier by Fowler and Nordheim. The third chapter gives some practical details of low pressure rectifiers, ignitrons and thyratrons.

Chapter IV deals with technical details of the glass bulb, steel tank and pumpless rectifiers and ignitrons, and discusses, in terms of the fundamental processes outlined in chapter II, some of the problems involved in the design and operation of these units. The latest type of high voltage ignitrons are described. The last chapter deals with some aspects of the current and voltage limitations and the life of rectifiers. An extensive bibliography is given.

It is clear from the treatment of this complex subject that the author appreciates the importance of a knowledge of fundamental processes of electrical discharges in gases for the understanding of rectifier problems, and recognizes the need for close co-operation between physicist and engineer in the design of rectifiers.

This nicely bound and well printed book of 239 pages will be of most use to those actively engaged in the design of rectifiers, but it will also be useful to engineers concerned with their operation and maintenance. C. GREY MORGAN.

An Introduction to Scientific Research, by E. BRIGHT WILSON. Pp. xiii + 375. (New York: McGraw-Hill, 1952.) 51*s*.

It is a far cry indeed from the classical scientific method, with its sterile discussion of Mill's canons, to Professor E. Bright Wilson's book in which the approach is avowedly practical and utilitarian. In his preface the author stresses that research is only too often " a highly inefficient endeavour ". Most of us will have to agree with him, if we are honest, but there will always be those who, conscious of the 'finger-sight' of the greatest experimentalists, will doubt whether books can be of very much help. Such a defeatist attitude is quite indefensible when we think in terms of the newcomer to research. There are tricks of the trade which the experienced can hand on to the apprentices. The sooner they are learnt the better. There is enough to acquire by the hard way of disappointment without refusing the very real assistance of this book.

It is impossible to write a book of this kind which will satisfy everyone. Argument as to selection of material and depth of treatment of the various topics is inevitable. Accepting the principle that the author's book which has been written is of greater interest than the reviewer's which has not, the fairest course is to indicate what ground the book covers.

A brief outline dealing with the choice and statement of research problems (with some pointed remarks about the cost of experiments) is followed by a good introduction to the use of literature, which could profitably be expanded for the English reader, and an outline of classical elementary scientific method.

An outstanding feature of the book is the stress laid on the contribution of statisticians to the development of an ordered body of principles for the design and analysis of experiments. This section of the book will prove of value even to those who have worked in research for many years without ever getting familiar with the ideas of randomization, replication, confounding, levels of significance and so on. They will find an easy introduction to sampling processes and the common sampling distributions such as the Binomial, Multinomial, Chi-square, Poisson and Gaussian, as well as Analysis of Variance in its simpler aspects.

There are two chapters dealing with mathematical methods (existence theorems, dimensional analysis, etc.) and numerical computation (nomographs, punched card computers, analogue machines, numerical solution of equations, etc.). Other chapters deal with the design of apparatus (servo systems, kinematic design, feedback, 'noise' limitations on measurements etc.), the execution of experiments, errors of measurement and reporting the results of the research.

The above very brief catalogue should indicate that the author has covered a very wide field in selecting his material and it is quite certain that workers in all fields will find a considerable amount that will repay careful study. It should not be counted as a drawback that an introductory book is only introductory. Professor Wilson achieves his avowed intention admirably and his book can be strongly recommended for all research libraries. M. J. M.

Designing by Photoelasticity, by R. B. HEYWOOD. Pp. xvi+414. (London: Chapman and Hall, 1952.) 65*s*.

Although many designers will disagree with Dr. Heywood's suggestion that the art of design consists mainly of an accurate determination of the stresses in a component, they should find the collection of information on stress-concentrations given in the second part of his book of considerable value in saving them the task of exploring a formidable number of publications in order to obtain the stress-concentration factor for a particular notch. Much of the information is presented in graphical form so that factors for almost any combination of notch dimensions can easily be found. In addition to the well known standard cases of grooves, shoulders and holes, the author gives information on the strength of ribbed plates in bending, and on the stresses in Tee headed members. Unfortunately the important cases of the splined shaft and key way have not been dealt with as thoroughly as other types of notch. A complete chapter has been given to the subject of screwed fasteners, and this should be useful to the designer who can afford the luxury of special thread forms and nuts in difficult cases. The final chapter could have been improved by a fuller treatment of the application of stress-concentration factors to cases of combined alternating stresses.

In the first part of the book the author gives a simplified account of the photoelastic effect and of some of the methods of evaluating stresses. He confines his attention almost exclusively to the determination of peak stresses on an unloaded boundary, which, as he points out, are the cause of failure in most engineering components. Unfortunately the methods given are insufficient for the exploration of stresses elsewhere, and the presentation of the theory is inadequate for the interpretation of those accidental optical and stress-optical effects which are liable to appear even in the simplest of cases. The author gives, however, much useful information on materials and techniques, and some illustrations, from his own work, of the application of photoelastic methods to the reduction of stress-concentrations. The book is well produced and fully illustrated and includes a comprehensive and well arranged bibliography.

H. T. JESSOP.
T. H. LAMBERT.

CORRIGENDUM

Interferometric Studies of the Growth of Stearic Acid Crystals and their Optical Properties, by AJIT RAM VERMA and P. M. REYNOLDS (*Proc. Phys. Soc.* B, 1953, **66**, 414).

P. 418. Table 1, column 5 (Multiple), item 2 *should read*:

$1\frac{1}{2}\times(43{\cdot}5\pm0{\cdot}6)$; $2\times(32{\cdot}6\pm0{\cdot}5)$

and item 3 *should read*:

$2\times(49{\cdot}3\pm1{\cdot}2)$; $2\frac{1}{2}\times(39{\cdot}5\pm1)$; $3\times(32{\cdot}9\pm0{\cdot}8)$

CONTENTS OF SECTION A

PAGE

Prof. Max Born. The Conceptual Situation in Physics and the Prospects of its Future Development. (37th Guthrie Lecture) 501

Dr. M. J. M. Bernal. Analytical Wave Functions for Methane and the Ammonium Ion 514

Mr. J. Kyles, Dr. C. G. Campbell and Dr. W. J. Henderson. An Investigation of the β-Particle Spectrum of Mesothorium 2 519

Dr. E. P. George, Mr. J. L. Redding and Dr. P. T. Trent. Cloud Chamber Observations of the Cosmic Radiation Underground 533

Mr. J. K. Parry, Dr. H. D. Rathgeber and Mr. J. L. Rouse. Ionization of Cosmic Ray Mesons in Argon 541

Dr. S. R. Haddara and Dr. D. Jakeman. The Lateral Structure of Cosmic Ray Extensive Air Showers at Sea Level 549

Mr. G. K. White. The Thermal Conductivity of Gold at Low Temperatures . 559

Dr. J. R. Holt and Mr. T. N. Marsham. An Investigation of (d, p) Stripping Reactions—IV : Results for ^{40}Ca and ^{88}Sr 565

Mr. B. V. Paranjape. On the Theory of Internal Friction in Metals 572

Research Notes :

Dr. P. G. Klemens. Electronic Thermal Conduction in Superconductors . 576

Miss D. Sayers and Mr. P. K. Carroll. The Wavelengths of Nitrogen First Positive Bands 577

Letters to the Editor :

Prof. E. W. Titterton and Mr. T. A. Brinkley. Cross Sections for the Reaction ^{7}Li(γT)^{4}He at 6·13, 14·8 and 17·6 mev 579

Reviews of Books 581

Contents of Section B 584

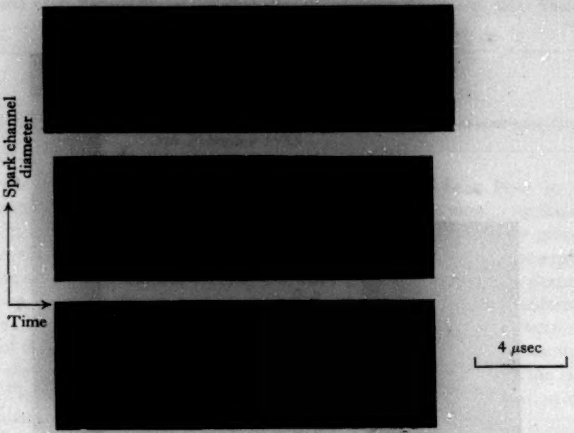

Spark channel diameter
Time
4 μsec

Fig. 1. Streak photographs of hydrogen spark channels with 1, 4 and 10 μsec rectangular current pulses.

Fig. 3. 4 μsec, 110 A peak, current pulse.

PLATE I

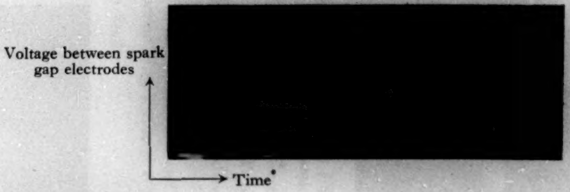


Fig. 4. Spark voltage oscillograms with 6 μsec pulses for gap lengths of 0·1, 1, 3, 5, 7 and 9 mm. The inflection at about 4 μsec is due to the shape of the current pulse (see fig. 3).

4102 Å
H_δ
4340 Å
H_γ
4861 Å
H_β
6563 Å
H_α

Fig. 10. Spectrum of 0·5 μsec hydrogen sparks.

Plate II

Heat Flow and the Growth of Metal Single Crystals from the Melt

By A. J. GOSS
The University, Southampton*

Communicated by S. Weintroub; MS. received 18*th November* 1952, *and in amended form* 5*th February* 1953

Abstract. A simple theoretical relation is derived for the heat flow in a crystallizing metal rod and three cases corresponding to slow, medium ($\simeq 5$ mm min^{-1}), and fast rates of growth, are considered. Details are given of temperature–time traces obtained experimentally using a thermocouple embedded in the solidifying metals, viz. Bi, Cd, In, Pb, Sn and Zn. The shapes of the traces are of three types corresponding to the three cases considered theoretically. It is shown that an approximate value of the ratio of the thermal conductivities of the solid and liquid metal can be obtained from the experimental results for the slow rate of growth, and values of the ratio are given for the six metals used. The importance of the heat flow on the form and orientation of the crystals is discussed.

§ 1. Introduction

The flow of heat from the crystallizing metal during the growth of metal single crystals from the melt has been the subject of several previous investigations. Goetz and Hasler (1929), Palibin and Froiman (1933), Chalmers (1949) and Gow and Chalmers (1951), for example, have discussed the heat flow problem qualitatively and have postulated various possible influences, for example the direction of the heat flow, on the growth. These influences have been supported to some extent by experiment. Theoretically, the heat flow problem has been tackled by, in particular, Roscoe (1934), Kuznecov and Saratovkin (1934) and Duran (1941). However, the assumptions made by both Kuznecov and Saratovkin, and Duran, were incomplete and their results are therefore of little practical value. The heat flow in a crystallizing rod was considered in detail by Roscoe and his results are of considerable interest. Nevertheless, Roscoe was unable to obtain satisfactory agreement between his theoretical and experimental values. More recently Pomeroy (private communication) has modified and improved Roscoe's treatment slightly but no experimental observations were made to test the revised relationship between the various factors controlling crystal growth. It would appear therefore that the various proposals regarding the influence of heat flow on crystal growth which have been put forward lack direct experimental confirmation.

In the following a fresh attack is made on the problem. A simple relation (eqn. (2)) for the heat flow in a crystallizing rod is given and this is used to distinguish three particular cases which are of practical importance. Experimental results, obtained in the form of temperature–time traces, on the crystallization of several metals are then discussed and it is shown that the traces are of three types in qualitative agreement with the above theory. The temperature of the solidifying metal was obtained by a thermocouple embedded in the metal and not, as in Roscoe's work, by a thermocouple placed outside the mould containing the metal.

---

* Now at the Bell Telephone Laboratories, Murray Hill, New Jersey, U.S.A.

The conditions which are assumed to apply in the derivation of eqn. (2) are very closely those which occur in practice for slow rates of growth, and, in this case, good quantitative agreement between the theoretical and experimental results is obtained. It is thus possible to calculate an approximate value ($\pm 10\%$) for the ratio of the thermal conductivities k_s and k_l, of the solid and liquid metal respectively. The experimental values which are obtained in this way for k_s/k_l for tin, lead, bismuth and zinc agree with those obtained by other methods by other investigators. For indium the value obtained, 2·3, is slightly greater than the corresponding ratio of the electrical conductivities; k_s/k_l for indium has not been determined previously. Finally, the connection between the heat flow and the form and orientation of the metal single crystals is discussed.

§ 2. THE HEAT FLOW AT CRYSTALLIZATION

Crystallization by slow solidification from the melt takes place by the transformation of the liquid metal into solid with the release of latent heat and this latent heat must be removed from the metal in order that the solidification may continue. Hence heat flow, which may take place by conduction, convection or radiation, away from the liquid–solid interface is essential for crystal growth. If the metal is *in vacuo* heat loss by convection may be disregarded.

It has been found experimentally that the temperature of the crystallizing metal in the slow solidification from the melt method may differ appreciably from that of its immediate surroundings. In some extreme cases it may differ by as much as 100°c above or below its surroundings. The thermal transfer by radiation is therefore a somewhat variable, and usually unknown, quantity. It is, however, possible to derive a simplified, and to some extent approximate, thermal relationship (eqn. (2)) which expresses the change in thermal conditions at solidification, and which is, in certain circumstances, in good accord with experiment.

For the heat flow across the solid–liquid interface of a crystallizing metal rod *in vacuo*, since in the limit when $dx \rightarrow 0$ only the thermal conduction across the interface and the latent heat evolved at the interface are involved (Roscoe 1934), it may be shown that

$$k_s A_s \left(\frac{dT}{dx}\right)_s - k_l A_l \left(\frac{dT}{dx}\right)_l = AL\rho\left(\frac{dx}{dt}\right) \qquad \ldots\ldots(1)$$

where the suffixes l and s denote the liquid and solid phases respectively, k, L, T and ρ the thermal conductivity, latent heat, temperature and density respectively of the metal, x the coordinate of the interface as measured along the rod, A the area of cross section of the rod and t the time. Only the thermal conduction across the interface and the latent heat evolved at the interface are involved. Although there is a small change in volume (about 3%) between the solid and liquid, we may consider $A_s \simeq A_l \simeq A$. Hence,

$$k_s G_s - k_l G_l = RL\rho \qquad \ldots\ldots(2)$$

where G and R denote, respectively, the temperature gradient, and the rate of movement of the interface. Three particular cases are worthy of special comment.

(*a*) *R small.* The latent heat term $RL\rho$ is then small and may be neglected, so that $k_s/k_l = G_s/G_l$. This applies whether k_s is greater or less than k_l.

(*b*) *Balanced condition of heat flow.* R is of such a value, depending on the thermal constants of the metal, that eqn. (2) holds for $G_l \simeq G_s$. This is referred to as the balanced condition of heat flow and can only occur when $k_s > k_l$.

(*c*) *R large.* The latent heat term $RL\rho$ is then large and this contribution to the heat flow can no longer be assumed to be removed only by conduction along the rod. A sharp solid–liquid interface and consequent neglect of the thermal transfer by radiation cannot then be assumed and eqn. (2) is no longer valid.

The importance of these three cases in relation to the experimental growth of crystals will be made clear in later sections. The above discussion is limited in application strictly to the central portion of a long rod. In addition, it is admittedly less comprehensive than Roscoe's (1934) treatment, but Roscoe introduces assumptions with regard to the heat transfer by radiation and to the temperature of the metal in the mould which cannot be justified experimentally.

§ 3. Experimental Procedure

Temperature–time traces obtained during the crystallization of tin, lead, zinc and bismuth have been described previously (Goss and Weintroub 1952). Additional experiments have since been made on cadmium, indium and zinc. The method used was described in detail by Goss (1952). It is essentially the Andrade and Roscoe (1937) horizontal travelling furnace method in which the specimens are crystallized by slow solidification from the melt. The single crystal rods were approximately 5 mm in diameter and 20–30 cm long. Batches, usually of four specimens, in separate 7–8 mm internal bore open-end Pyrex moulds, were placed at the middle of the furnace tube. The pressure inside the tube was about 0·01 mm Hg. The furnace was moved at a steady rate and passed completely over the specimens so that crystallization commenced at one end and continued uniformly along the specimens. The variation in temperature of the metal during the run was recorded by means of a chromel–alumel thermocouple embedded in one of the specimens. The thermocouple consisted of silver-soldered wires. Both 33 and 40 swg wires were used but in spite of the difference in thermal capacity of the corresponding junctions the temperature records in the two cases were identical.

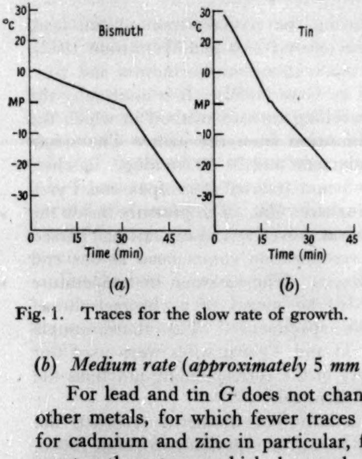
The thermal e.m.f. was recorded on a Kent potentiometric recorder, the scale of which was 250 mm wide, reading to 6 mv, corresponding to about 150°c. From the recorder trace which represents the time variation of the temperature of the mid-point of a specimen, the temperature gradient, at any time, along the rod can be determined, for approximately uniform conditions over the central length of the rod. A delay between the recorded e.m.f. and the actual thermocouple e.m.f. is inherent in the balancing system of the potentiometer (except when the thermal e.m.f. is constant), but for a steady rate of change of temperature this delay is constant and thus no appreciable error enters into the measurement of the rate of change of temperature. The correctness of the traces was confirmed by independent measurements with a direct reading potentiometer. In addition, to ascertain the direction of the heat exchange between the specimens and their surroundings, measurements were made with differential thermocouples and pairs of thermocouples. It was shown also that the rate of crystal growth was identical with the rate of furnace travel. The connection between the form of the crystals and the rates of growth has been discussed elsewhere (Goss and Weintroub 1952). The relationship between the temperature traces, the heat flow and crystal form only will be considered here.

§ 4. THE TRACES

There are three types of temperature–time traces corresponding to the three rates of growth (*a*) slow, (*b*) medium and (*c*) fast.

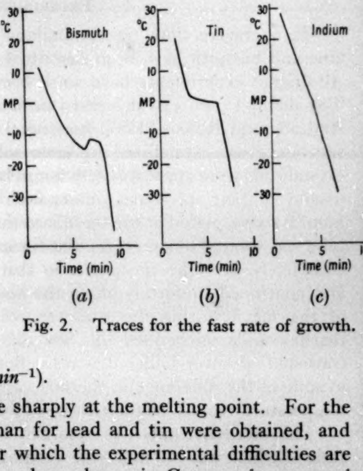
(*a*) *Slow rates* (1 *mm* min^{-1} *or less*) (see fig. 1).

There is a characteristic change in slope as the metal cools through the melting point. For cadmium, indium, lead, tin and zinc, G_s is less than G_l. The gradients just above and below the melting point are quite uniform but the gradient changes sharply at the melting point. For bismuth G_s is greater than G_l. For the slow rates the traces obtained are in agreement with § 2, case (*a*); the numerical values of G are considered later (see § 5).



(*a*) (*b*)

Fig. 1. Traces for the slow rate of growth.



(*a*) (*b*) (*c*)

Fig. 2. Traces for the fast rate of growth.

(*b*) *Medium rate* (*approximately* 5 *mm* min^{-1}).

For lead and tin G does not change sharply at the melting point. For the other metals, for which fewer traces than for lead and tin were obtained, and for cadmium and zinc in particular, for which the experimental difficulties are greater, those traces which do not show a sharp change in G cannot be accepted without further qualification.

It may be observed from eqn. (2) that the balanced condition of heat flow is attained when the change in thermal conductivity, k_l to k_s, exactly compensates for the release of latent heat. Equal temperature gradients in the liquid and solid at solidification are then possible. For lead and $G = 20°\mathrm{c\,cm}^{-1}$, we find from eqn. (2), $R = 6\,\mathrm{mm\,min}^{-1}$, and for tin and $G = 15°\mathrm{c\,cm}^{-1}$, $R = 5\,\mathrm{mm\,min}^{-1}$. These values of R are the rates at which it is found experimentally that lineage (as defined by Goss and Weintroub 1952) is greatest. Lineage was observed also for bismuth (for slow rates), but not for cadmium, indium and zinc.

(*c*) ***Fast rates*** **(10** ***mm*** $\boldsymbol{min^{-1}}$ ***and greater*****).**

Here G_s is greater than G_l. For bismuth $G_s - G_l$ is much greater than for the slow rates (compare the traces for bismuth in figs. 1 (*a*) and 2 (*a*)) and supercooling occurs nearly always. For the other metals, for which k_s is greater than k_l, the traces assume two forms:

(A) as for lead and tin, at the melting point the temperature remains constant, in some cases for several minutes, followed by a sudden drop (fig. 2 (*b*)); under these conditions single crystals are formed.

(B) as for cadmium, indium and zinc, there is only a sharp increase in G at solidification (fig. 2 (*c*)); single crystals are not then obtained.

It is not possible to correlate the values of G, obtained from the traces for the fast rates, with eqn. (2), because there are large temperature differences between the specimen and the surroundings and therefore large radiation effects.

§ 5. The Ratio of the Thermal Conductivities of the Solid and Liquid Metal

Equation (2) may be rewritten in the form

$$G_s/R = G_l k_l/Rk_s + L\rho/k_s \qquad \ldots\ldots(3)$$

showing that G_s/R is linearly related to G_l/R. For the six metals data obtained from the temperature–time traces for the slow rates were used to calculate the values of G/R. These values were plotted (not shown) and the linear relationship was confirmed. As expected, since the radiation from the interface cannot be neglected, G_s/R and G_l/R for the fast rates were not linearly related. The linear graph emphasizes unduly the values obtained with very small values of R and eqn. (3) was therefore rewritten in the form

$$G_s - RL\rho/k_s = G_l k_l/k_s. \qquad \ldots\ldots(4)$$

$L\rho/k_s$ was calculated for each metal, using accepted values, and the left-hand side of eqn. (4) was plotted against the corresponding value of G_l (see figs. 3 and 4). An approximate value of k_s/k_l was obtained from the slope of the lines drawn through the points. In the table these ratios are compared with those of previous investigators.

Experimental Values of k_s/k_l for Various Metals

Bismuth	Cadmium	Indium	Lead	Tin	Zinc
0·45 (*a*)	2·4 (*b*)		1·83 (*a*)	1·7 (*a*)	1·51 (*a*)
0·65 (*c*)			1·24 (*d*)	1·7 (*b*)	1·56 (*e*)
				2·0 (*c*)	
0·57 (*f*)	1·6 (*f*)	2·3 (*f*)	1·7 (*f*)	1·9 (*f*)	1·8 (*f*)

(*a*) Konno 1919, (*b*) Brown 1923, (*c*) Northrup and Pratt 1917, (*d*) Bidwell 1940, (*e*) Bidwell 1939, (*f*) present work. The values quoted are subject to a variation of not more than ± 10%.

The agreement for bismuth, lead, tin and zinc is satisfactory. The values of k_s of the previous investigations are those for polycrystalline material whereas those for the present work are for single crystals. However, the difference between k_s for a polycrystal and for a single crystal of the usual orientation is very small. For cadmium the value of $k_s/k_l = 1{\cdot}6$ is considerably lower than 1·9, the corresponding electrical ratio, and is very different from Brown's value, 2·4, which is considered to be unusually large (Powell 1949). A value of about 1·8 is probably correct. For indium the value 2·3 obtained for k_s/k_l is the only one known. It is greater than the value, 1·9 approximately, of the corresponding electrical ratio, but the experimental evidence, given by the several temperature–time traces obtained for indium, shows that 2·1 is the least possible value for the thermal ratio.

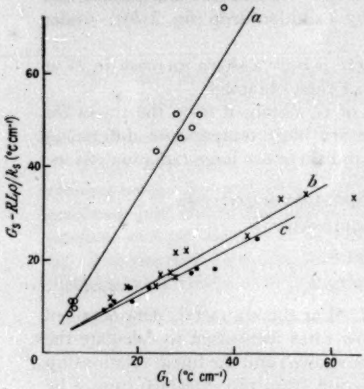


Fig. 3. Values of eqn. (4), (*a*) bismuth, (*b*) lead and (*c*) tin.

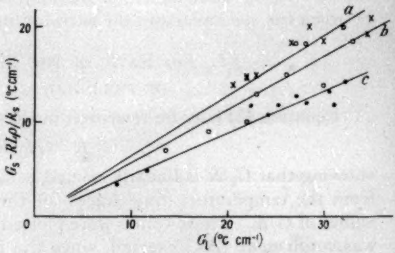


Fig. 4. Values of eqn. (4), (*a*) cadmium, (*b*) zinc and (*c*) indium.

§ 6. THE INFLUENCE OF HEAT FLOW

It is worth while considering whether there is any connection between the resultant form of the crystals grown and the heat flow during growth. In § 4 (*b*) it was mentioned that lineage is most strongly developed in lead and tin crystals grown at rates which correspond to 'balanced' heat flow. The evidence of other workers on this point is however somewhat contradictory. Puttick and King (1949), for example, state that tin crystals grown in graphite moulds show less tendency for lineage than when grown in moulds of materials of very poor thermal conductivity. However, more recently, Teghtsoonian and Chalmers (1951), have observed lineage in tin crystals grown at rates of 1–10 mm min^{-1} in graphite moulds. This is to be expected since the heat developed at solidification is removed from the interface to a considerable extent by conduction through the mould and the heat flow is more nearly in the 'balanced' condition.

For bismuth, since $k_s < k_l$, the 'balanced' condition of heat flow cannot be attained. However, lineage did occur in the crystals grown at the slowest rates. The cadmium, zinc and indium crystals were free from lineage, though it should be mentioned in this connection that only a few specimens of cadmium or indium grown at the medium rate were examined.

For the fast rates the contrast in crystallization between lead and tin on the one hand, and cadmium, indium and zinc on the other, is very marked. This is in accordance with the two types of temperature traces obtained for the fast rates (fig. 2). For lead and tin the records show a steady temperature at the melting point and single crystals were obtained, but for the other metals this steady temperature was absent and only strips of crystals, of dissimilar orientations, were formed.

§ 7. PREFERRED ORIENTATION AND THERMAL CONDUCTIVITY

Several tentative theories to account for the presence of preferred orientation in crystals grown from the melt have been advanced, but up to now the experimental data to test these theories have by no means been conclusive. It is now possible to correlate the orientation with the physical properties, particularly the thermal conductivity, of the crystal.

For slow rates of growth, bismuth, cadmium and tin of the metals studied showed preferred orientation, but indium and zinc did not. The apparent preference for orientations of large ψ, the angle between the specimen and a crystallographic axis, has already been the subject of comment (Goss 1951). At slow rates the heat flow is from the crystallizing interface along the specimen, the temperature of which is about the same or lower than that of its surroundings, and therefore the crystal will grow with a preference for that orientation for which the heat is conducted away most readily. For bismuth, cadmium and tin the thermal conductivity is greatest in the direction perpendicular to the axis of three-, six- and four-fold symmetry respectively, i.e. in the basal or cleavage plane, and is approximately 1·3 times the conductivity in the axial direction. For zinc the thermal conductivity changes very little with orientation (ratio of 1 : 1·05), and no preferred orientation occurs at slow rates of growth. No particular preferred orientation has been observed for indium. This is in agreement with some recent measurements by the author on indium single crystals which indicate that the electrical resistance, and hence the thermal resistance, is approximately isotropic (ratio of $<1:1{\cdot}05$).

The results may be different, however, when stress is present; for example, zinc, grown in a vertical position, has been found to show a preference for its *c*-axis to be perpendicular to the specimen axis (Kuznecov and Saratovkin 1934). The orientating effect of heat flow is also clearly observed in the work of Goetz and Hasler (1929), Palibin and Froiman (1933) and Slifkin (1951), but it is possible that some degree of constraint may have been present in all these experiments.

For fast rates of growth, and using the horizontal method, the orientations which result are not necessarily the same as those for slow rates. Zinc is an extreme case; at fast rates, there is a very strong preference for the hexagonal axis to be normal to the specimen axis. Bismuth retains its preference for large values of ψ between the trigonal and specimen axes, but tin no longer shows a preferred orientation. Since the heat flow, at fast rates of growth, is governed by radiation in addition to conduction, there is little thermal control of the orientation of the crystals. The metals that exhibit preferred orientation at the fast rates crystallize with orientations in the direction of least compressibility and therefore of greatest binding. Tin, however, is approximately isotropic in this respect.

§ 8. CONCLUSION

It is apparent therefore, from the observations on the various metals, that no one factor is responsible for the preferred orientations which occur in the growth of single crystals. It is considered, however, that at the slow rates of growth it is the thermal conductivities which are mainly responsible, but that other factors are of great importance when rapid or stressed growth occurs.

ACKNOWLEDGMENTS

The author wishes to thank Professor A. M. Taylor and the University of Southampton for facilities, the Ministry of Education for a Further Education and Training Grant, and Mr. S. Weintroub for his encouragement and advice. This work forms part of a thesis accepted by the University of London for a Ph.D. Degree.

REFERENCES

ANDRADE, E. N. DA C., and ROSCOE, R., 1937, *Proc. Phys. Soc.*, **49**, 152.
BIDWELL, C. C., 1939, *Phys. Rev.*, **56**, 594; 1940, *Ibid.*, **58**, 561.
BROWN, W. B., 1923, *Phys. Rev.*, **22**, 171.
CHALMERS, B., 1949, *Proc. Roy. Soc.* A, **196**, 64.
DURAN, A., 1941, *Anal. Fis. y Quim.*, **37**, Supp. 33.
GOETZ, A., and HASLER, M. F., 1929, *Proc. Nat. Acad. Sci., Wash.*, **15**, 646.
GOSS, A. J., 1951, *Research, Lond.*, **4**, 292; 1952, *Thesis*, University of London.
GOSS, A. J., and WEINTROUB, S., 1952, *Proc. Phys. Soc.* B, **65**, 561.
GOW, K. V., and CHALMERS, B., 1951, *Brit. J. Appl. Phys.*, **2**, 300.
KONNO, S., 1919, *Sci. Rep. Tohoku Imp. Univ.*, **8**, 169.
KUZNECOV (Kuznekov)*, V., and SARATOVKIN, D., 1934, *C. R. Acad. Sci., U.R.S.S.* (new series) **1**, 248.
NORTHRUP, E. F., and PRATT, F. R., 1917, *J. Franklin Inst.*, **184**, 675.
PALIBIN, P. A., and FROIMAN, A. I., 1933, *Z. Kristallogr.*, **85**, 322.
POWELL, R. W., 1949, *J. Iron Steel Inst.*, **162**, 315.
PUTTICK, K. E., and KING, R., 1949, *Royal Aircraft Establishment*, Report No. Met. 45.
ROSCOE, R., 1934, *Thesis*, University of London.
SLIFKIN, L. M., 1951, *J. Appl. Phys.*, **22**, 1216.
TEGHTSOONIAN, E., and CHALMERS, B., 1951, *Canad. J. Phys.* **29**, 370.

* Transliteration of Russian name according to the Royal Society's scheme *Transliteration of Russian, Serbian and Bulgarian for Bibliographical Purposes.*

Electron Currents in Thin Oxide Films on Aluminium

By A. CHARLESBY

Atomic Energy Research Establishment, Harwell, Didcot, Berks.

MS. received 27th January 1953

Abstract. At electric fields lower than those required to cause an appreciable ion current to flow, an electron current is observed in thin films of oxide formed on aluminium in suitable electrolytes. The relationship between electron current i_- and the field F across the film is represented by the equation $i_- = A_- \sinh B_- F$ where A_-, B_- depend on the temperature and on the barrier dimensions opposing the current. These latter are found to be about 0·61 ev and 1·4 Å. A similar formula also holds for tantalum. Taken together with the equation for ion current published previously, this equation represents the entire current over a range of some 10^4 to 1.

The presence of trapped electron charges in the oxide layer near the electrolyte surface is discussed.

§ 1. Introduction

When aluminium is anodized in suitable electrolytes, a thin insulating oxide film is formed which acts as the dielectric of a condenser. During the formation process the current flowing consists mainly of ions, which build up the film. In a previous paper (Charlesby 1953) the relationships between this ionic current, the formation voltage, film thickness and temperature were studied. If the voltage applied to a film is reduced well below the formation voltage, the current drops markedly and no longer results in film growth. Even after a long period, when a considerable total charge has passed, there is no appreciable increase in the thickness of the oxide layer. It is therefore necessary to assume that at these low current densities the current consists mainly or entirely of electrons, which can flow without affecting the oxide film. This electron current, which can flow indefinitely provided that a constant voltage is maintained across the film, is often referred to in the literature as the 'leakage current'. In this paper the relationship between electron current i_-, voltage across the film V, film thickness δ and temperature T is studied. In addition to our own results, some experimental data obtained by Guntherschulze and Betz for aluminium and tantalum (which behaves in an analogous fashion) have been analysed.

§ 2. Theoretical Relationship between Electron Current, Field and Temperature

The oxide film formed on aluminium by anodizing is of uniform thickness, since slight variation would cause small changes in the field, and considerable ones in the ion current which builds up the film. Only the electron movement across the film, at right angles to its surface, need therefore be considered. It is assumed that this flow of electrons is inhibited by a potential barrier of height U and half width a. This barrier may be considered to occur at or near the metal–oxide surface, although this assumption is not essential. In the absence of an external field F, the probability of an electron jumping the barrier is $\nu \exp(-U/\boldsymbol{k}T)$ where ν is the probability (per second) of a transition in the absence of the barrier.

If a field F (assumed uniform) is present across the film, at right angles to its surface, electron motion in the direction of the field will be aided, while in the opposite direction it will be opposed. The effect is to replace the barrier height in one direction by $U-a\boldsymbol{e}F$, and in the other by $U+a\boldsymbol{e}F$, $\boldsymbol{e}$ being the electronic charge. The flow of electrons will then be

$$\begin{aligned} i_- &= N\boldsymbol{e}\nu \exp(-U/\boldsymbol{k}T)\{\exp(a\boldsymbol{e}F/\boldsymbol{k}T)-\exp(-a\boldsymbol{e}F/\boldsymbol{k}T)\} \\ &= 2N\boldsymbol{e}\nu \exp(-U/\boldsymbol{k}T)\sinh(a\boldsymbol{e}F/\boldsymbol{k}T) \qquad \ldots\ldots(1) \end{aligned}$$

where N electrons are available to carry the current. At constant temperature (1) can be written in the form

$$i_- = A_- \sinh B_- F = A_- \sinh V/V_1 \qquad \ldots\ldots(2)$$

where $A_- = 2N\boldsymbol{e}\nu \exp(-U/\boldsymbol{k}T)$, $B_- = a\boldsymbol{e}/\boldsymbol{k}T$ and $V_1 = \delta/B_-$. Here the field F is taken as V/δ, the potential V across the oxide film divided by its thickness δ. At low fields ($a\boldsymbol{e}F/\boldsymbol{k}T \ll 0{\cdot}8$)

$$i_- = A_- B_- V/\delta \qquad \ldots\ldots(3)$$

while at high fields and in the absence of ion current

$$\log i_- = 0{\cdot}4343 B_- V/\delta + \log(A_-/2). \qquad \ldots\ldots(4)$$

In what follows these theoretical relationships are compared with experimental determinations of the current–voltage curves. In addition to verifying eqn. (1), the comparison gives values for A_- and B_-, and hence U and a.

§ 3. VARIATION OF ELECTRON CURRENT WITH APPLIED VOLTAGE AT CONSTANT TEMPERATURE

It is convenient to plot the experimental current–voltage curves on double logarithmic paper and compare the resultant plot with the curve $y = \sinh x$ drawn to the same scale. If the theoretical relation (2) is correct, the two curves should coincide, apart from horizontal or vertical shifts of the axes. These shifts give the magnitude of A_- and V_1, which are the current and voltage corresponding to $y=1$, $x=1$. If V_f is the formation voltage, the film thickness δ is approximately $1{\cdot}1 \times 10^{-7} V_f$ at room temperature. Then an approximate value for B_- can be deduced: $B_- \sim 1{\cdot}1 \times 10^{-7} V_f/V_1$. A comparison of the observed current–voltage relationship with formula (2) is shown in fig. 1. Good agreement is obtained except at voltages close to the formation voltage, when the ion current is no longer negligible.

Experimental values of A_-, V_1 and V_f/V_1 for a number of specimens are given in table 1. Some variation in these values is possible without seriously affecting the goodness of fit. A further cause of error is due to variation in the ratio δ/V_f. This depends to some extent on the formation current, and by assuming a constant value at room temperature (20°C) errors of 20% may arise in the value of B_-.

In a well formed film and at low fields ($V_f < 0{\cdot}1$ v) considerable hysteresis effects are observed when the voltage is changed. As a result it is difficult to obtain reproducible results at these low voltages, the electron current depending on the time spent at a given voltage and on whether this had previously been increased or reduced. These hysteresis effects are ascribed to the formation of space charges within the oxide layer, due to the trapping of electrons. The existence of these space charges has been demonstrated experimentally.

Table 1. Electron Current in Relation to Formation Voltage (20°c)

$$i = A_{-} \sinh V/V_1$$

Specimen	V_f (v)	A_{-} (μA cm^{-2})	V_1 (v)	V_f/V_1
I	120	0·01	23	5
III	330	0·01	67	5
IV	330	0·05	63	5
IV	260	0·017	45	6
	260	0·006	33	8
VII	50	0·0003	6	8
	81	0·006	15	5
	139	0·013	20	7
VIII	193	0·007	27	7
N	95	0·001	20	5
	95	0·0018	25	4
B	15	0·007	2·5	6
	80	0·04	20	4
	115	0·05	24	5
	148	0·075	31	5
H I	102·4		27	4
H10	12·75	0·006	3·3	4
R2	49	0·0003	10	5
		0·0004	12	4
	49	0·0002	13	4
	109	0·0008	30	4
Electrolytic condenser				
E1B/15	50	0·003	13	4

According to formula (3) $i_{-} \propto V$ at low field intensities. This ignores the field due to space charges within the film. It can be shown that, under equilibrium conditions, the effect of space charge in the neighbourhood of an electron emitter in vacuum is to reduce the effective field, so that $i \propto V^n$ where n equals 1·5. In the present experiments it was not feasible to wait for equilibrium to be reached. The longer the time lapse after a voltage charge, the smaller was the difference in electron current as between measurements with rising voltage and falling voltage. Moreover the average of these two values was approximately independent of this time lapse. Accordingly electron currents were measured 30, 45 and 60 seconds after each change in voltage, for both rising and falling voltage, and averaged. In this way an approximate value for the equilibrium current at any given voltage was obtained.

At *low* field intensities it was found that $\log i \propto n \log V$ where n varies between 0·7 and 2·0. Guntherschulze and Betz (1934a) obtained values of n between 1·966 and 2·99, depending on whether the field was increased or decreased, and on the time allowed for the current to settle down. In the literature they also give other and often incompatible relationships.

At *intermediate* values of V excellent agreement was observed between the predicted and observed shape of the $(\log i, \log V)$ curve.

At *high* values of V, approaching the formation voltage, the current increases more rapidly than is shown in the theoretical formula, the difference being due to the ion current. When this is allowed for, good agreement results, as is shown below for specimens of aluminium and tantalum.

Experimental Value of B_-

In table 1 the values of V_1, V_f and V_f/V_1 corresponding to the best fit with the theoretical curve are given for a number of aluminium specimens. Over a wide range of specimens and formation voltages it is found that V_f/V_1 is of the order of 5. Variations between one specimen and another may be largely ascribed to variations in the ratio δ/V_f.

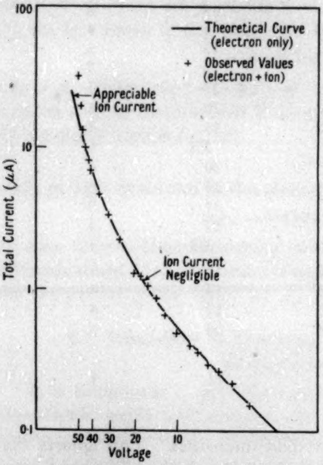


Fig. 1. Current–voltage curve for aluminium specimen formed for 10 hours at 20°c ($V_f = 50$ v).

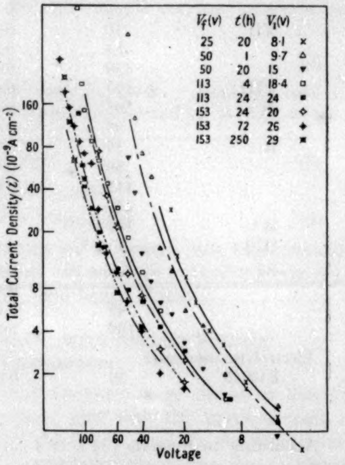


Fig. 2. Electron current for various film thicknesses (aluminium). (Curves represent theoretical formula $i = A \sinh V/V_1$.)

Assuming a value of $\delta/V_f = 1{\cdot}1 \times 10^{-7}$ for a well formed specimen, $B_- = 1{\cdot}1 \times 10^{-7} V_f/V_1 = 5{\cdot}5 \times 10^{-7}$. From formula (2), with $\mathbf{e} = 1$ electron charge, we obtain $a \simeq 1{\cdot}4$ Å.

From some experimental data given by Guntherschulze and Betz we deduce

$B_- = 8{\cdot}8 \times 10^{-7}$; $a = 2$ Å (1934 a)
$B_- = 7{\cdot}11 \times 10^{-7}$; $a = 1{\cdot}7$ Å (1934 b, p. 369)
$B_- = 5{\cdot}3 \times 10^{-7}$; $a = 1{\cdot}2$ Å (1934 a, fig. 4)

An extended series of investigations with an electrolytic condenser (the electrolyte consisting of glycol and ammonium borate), in which the film thickness was deduced from capacity measurements, gave the following values at 19°c: $V_1 = 15{\cdot}7$ v; $\delta = 690$ Å; $B_- = 4{\cdot}4 \times 10^{-7}$; $a = 1{\cdot}1$ Å.

Experimental Value of A_-

At room temperature the value of A_- varies irregularly from one specimen to another. It appears to depend on impurities present as well as on the time of formation. Values for A_-, deduced from the results of Guntherschulze and Betz, average around $1\,\mu\text{A}/\text{cm}^2$. By prolonged formation and high purity of electrolyte we have obtained much lower values (table 1). For any given specimen the value of A_- does not vary appreciably with film thickness. In fig. 2 the current–voltage relationship is shown for a single specimen formed at various voltages, and for varying lengths of time. With increasing time of formation voltage, and hence film thickness, the curve is displaced towards higher voltages V_1, due to increased film thickness, but the value of A_- remains unaltered to within experimental error. A similar conclusion was reached for the leakage current through an electrolytic condenser, A_- being independent of film thickness or film age when the formation voltage or time of formation was increased.

§ 4. VARIATION OF ELECTRON CURRENT WITH TEMPERATURE

At constant field the electron current increases rapidly with temperature. The empirical formula quoted in the literature is $\log i_- = \alpha + \beta T$. Values for α and β read from tables in the literature or abstracted from graphs are as shown in table 2.

Table 2. Variation of Electron Current with Temperature

Metal	α	β	Remarks	References
Aluminium	$\bar{9}\cdot13$	0·0236	$V_f = 300$, $V = 200$	GB 1934 a, p. 94
Aluminium		0·026		H 1940.
Tantalum	$\bar{9}\cdot8$	0·034	Temperature rising	GB 1932, p. 595
	$\bar{9}\cdot5$	0·029	Temperature falling, thicker film.	
Tantalum	$\overline{12}\cdot79$	0·0243	$V_f = 150$, $V = 50$	GB 1934 a, p. 94
Titanium	$\bar{8}\cdot7$	0·0237	$V_f = 14$, $V = 14$	GB 1934 a, p. 94

GB = Guntherschulze and Betz ; H = Herrmann.

The value of α varies with the voltage applied, and has no fundamental significance. In every case it is essential to keep the reference voltage well below the formation voltage, to ensure that ion current does not flow with increasing temperature. The similarity in the values of β for aluminium, tantalum and titanium is to be noted. Over a limited temperature range formula (1) can be written

$$\Delta \log i_- = 0\cdot4343 \frac{U - a\boldsymbol{e}F}{\boldsymbol{k}T^2} \Delta T$$

provided that $a\boldsymbol{e}F/\boldsymbol{k}T$ is greater than about 1·5. Comparison with the empirical formula $\Delta \log i = \beta \Delta T$ gives for

$$\beta = 0\cdot024, \quad U - a\boldsymbol{e}F \sim 0\cdot45 \text{ ev}, \quad U \sim 0\cdot5 \text{ ev},$$
$$\beta = 0\cdot028, \quad U - a\boldsymbol{e}F \sim 0\cdot52 \text{ ev}, \quad U \sim 0\cdot6 \text{ ev}.$$

A more exact determination is not justified since the exact values of F and T at which the measurements were taken are not available.

A detailed investigation on the variation of electron current with temperature has been carried out, use being made of an electrolytic condenser formed at 50 v, 100°c. The current–voltage relationships measured at varying temperatures are shown in fig. 3 and are compared with the theoretical formula (2). At 100°c the best fit is obtained for $V_1=20$ volts. The corresponding values of V_1 at other temperatures can be deduced since, according to theory, $V_1 \propto 1/B \propto T$. An experimental value of A_- (A_{obs}) at each of these temperatures can then be obtained by fitting the theoretical curve to the observed points, with the appropriate calculated value of V_1. Again a semi-theoretical value of A_- (A_{calc}) can be calculated for the same temperature from (2) assuming $U=0{\cdot}61$ ev, $2N\boldsymbol{e}\nu=4\times10^3$ A cm^{-2}. Table 3 shows the close agreement obtained between the calculated and observed values of A_- over a range of temperatures. Agreement as far as the variation of A_- and B_- with temperature, the value of U, and the general shape of the curve is within experimental error.

Table 3. Observed and Calculated Parameter A_- in Formula for Electron Current

T (°c)	100	80	65	50	37
V_1 (v)	20	18·8	18·1	17·3	16·6
A_{obs} (μA cm^{-2})	25	10	3·2	1·3	0·6
A_{calc} (μA cm^{-2})	25	9	3·5	1·3	0·54

§ 5. COMPLETE CURRENT–VOLTAGE CURVE

Having established the separate laws relating ion and electron current to applied field, it is interesting to study the extent of agreement for the complete current–voltage curve taken at various temperatures.

The formula for the total current i can be written as follows:

$$i=A_+\exp(B_+F)+2N\boldsymbol{e}\nu\exp(-U/\boldsymbol{k}T)\sinh(a\boldsymbol{e}F/\boldsymbol{k}T). \qquad ......(5)$$

The parameters A_+, B_+ refer to the ion current, which is only appreciable at voltages V greater than about 80% of the formation voltage V_f. The variation of A_+ and B_+ with temperature can be calculated if the barrier height opposing ion flow is known. If $V<0{\cdot}8V_f$ only three parameters, $N\boldsymbol{e}\nu$, U and a (the barrier dimensions limiting the electron current), need be considered.

Aluminium. Figure 3 shows the total current through an electrolytic condenser at various voltages and temperatures. The various symbols represent observed values, while the broken line represents the calculated relationship, with parameters for the electron current, as given above: $a=1{\cdot}1$ Å, $U=0{\cdot}61$ ev, $2N\boldsymbol{e}\nu=4\times10^3$ A cm^{-2}. The full line represents the same relationship, but with $a\sim1{\cdot}3$ Å. For voltages below about 45 and temperatures between 37°c and 100°c good agreement is obtained with the theoretical curves over a range of 1000 to 1 in current density. Above this voltage there is the expected steep rise due to the intrusion of an ionic current. The very small voltage range prevents a detailed quantitative comparison of the latter with theory. At lower temperatures and fields agreement is less satisfactory, due to an appreciable space charge effect.

Tantalum. The theoretical expression for the total current (5) is compared with some published data for a tantalum oxide film (Guntherschulze and Betz 1931, p. 156). The range of currents studied varies over a range of about 20000 to 1. The curve in fig. 4 shows more clearly than that for aluminium oxide the distinction between electron current and ion current. Below 80 volts the current is almost entirely due to electrons; above 100 volts only the ionic current need be considered.

As measurements were only made at one temperature, it is not possible to determine separately the values of $2Ne\nu$ and U from this data. However, assuming a value of 0·5 to 0·6 ev for the barrier height opposing electron flow, the following values of the parameters give close agreement with the published data (represented by crosses in fig. 4). Two values for ion current parameters are also given.

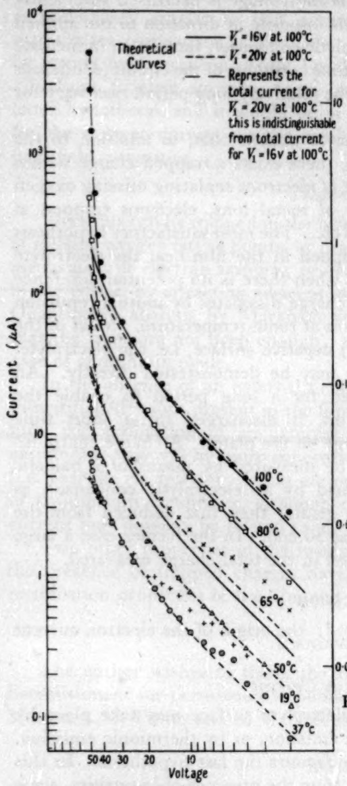


Fig. 3. Complete current–voltage curves at various temperatures (aluminium specimen formed at 50 v; 100°c, surface area 230 cm²).

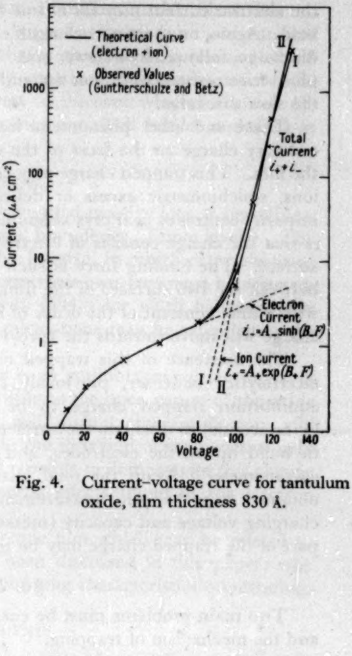


Fig. 4. Current–voltage curve for tantulum oxide, film thickness 830 Å.

Electron current $a = 0{\cdot}41$ Å; $u = 0{\cdot}5$ to $0{\cdot}6$ ev;

$$2Ne\nu = 355 \text{ to } 2 \times 10^5 \text{ A cm}^{-2}.$$

Ion current $A_+ = 6{\cdot}7 \times 10^{-21}$ (I) to $4{\cdot}26 \times 10^{-22}$ A cm^{-2} (II);

$$B_+ = 1{\cdot}7 \times 10^{-6} \text{ (I) to } 1{\cdot}888 \times 10^{-6} \text{ cm v}^{-1} \text{ (II)}.$$

Space Charge Effects

A number of phenomena render it necessary to elaborate the simple picture of the oxide film as consisting of an insulator, on each face of which charges may be placed. Among these phenomena the most striking is the hysteresis which occurs in the electron current when the applied field is changed. If, for example, the potential across the film is reduced, several minutes elapse before the electron current approaches a new steady value. If the voltage is decreased sufficiently the electron current may for a time become opposite in direction to the applied field. Again, on shorting a charged electrolytic condenser, there is an immediate discharge, followed by a slower one. The time constants of the circuit (condenser plus series resistance) cannot account for the relatively long period necessary for the slow discharge.

These and other phenomena lead one to assume that, in addition to the ordinary charge on the faces of the oxide, there exists a trapped charge within the film. This trapped charge may consist of electrons replacing missing oxygen ions, stoichiometric excess or deficiency of metal ions, electrons trapped at impurity centres or near crystal boundaries etc. The most satisfactory hypothesis is that the charge consists of electrons trapped in the film near the electrolyte surface. The binding force is such that, when there is no potential difference between the two surfaces of the film, the charge dissipates by mutual repulsion with a time constant of the order of minutes at room temperature. Most of the charge will move towards the (previously) negative surface, i.e. the electrolyte.

The existence of this trapped charge may be demonstrated directly. An electrolytic condenser, previously charged for a long period to enable the equilibrium trapped charge to be set up, is discharged for a short time (1–5 seconds) to dissipate the surface charge on the plates. A charge continues to build up on the electrodes, and can be measured by means of a ballistic galvanometer. Again the total charge held by an electrolytic condenser, as obtained from ballistic measurements, is greater than that deduced from the charging voltage and capacity (measured at 50 c/s). In the former case a large part of the trapped charge may be included in the total charge measured.

§ 6. DISCUSSION

Two main problems must be considered: the origin of the electron current and the mechanism of trapping.

Origin of Electron Current

The emission of electrons from the electrolyte surface may take place by the quantum tunnel effect, by strong field emission, or by thermionic emission. The variation of current with temperature favours the last hypothesis. In this case the electrons presumably arise mainly from the presence of impurities, since there is evidence that these determine the magnitude of electron current. With very pure specimens the electron flow, at a given field intensity, is much lower than for less pure specimens.

Intrinsic conduction by the film cannot account for the observed electron currents. For such conduction the conductivity σ is of the form $\sigma = \sigma_0 \exp(-U/\boldsymbol{k}T)$ with $\sigma \sim 100\,\text{ohm}^{-1}\,\text{cm}^{-1}$. In our case σ_0 is of the order of $10^{-4}\,\text{ohm}^{-1}\,\text{cm}^{-1}$.

Once emitted, the flow of electrons to the anode is determined by the field in the manner studied quantitatively above. Assuming that the effective field is V/δ, then the barrier dimensions are 1·1–1·4 Å width (trough to peak) and 0·61 ev height. However, the presence of trapped charges may reduce the effective field F' near the electrolyte to a lower value. This will not affect the agreement with experimental data found in this paper, but the value of the barrier width will be increased.

Similar electron currents are observed for other insulators at fields of the order of 10^6 v cm^{-1}. Indeed the general shape of the (i, V) curve for such insulators is remarkably similar to that observed for aluminium oxide films, but no quantitative comparison is available. Von Hippel accounts for these currents by strong field emission, while Frenkel believes that the field assists thermal vibrations to eject electrons from impurity centres. Our results favour the latter hypothesis, and it would be of interest to discover whether formula (1) for the electron current fits the results obtained for alkali halides and mica at high field intensities.

Trapped Charges

Electrons can be trapped in the field of interstitial positive ions at impurities, at missing oxygen lattice points, or near crystal bounderies (Tamm levels). The mechanism of electron trapping proposed by Landau, in which extra electrons in the lattice, by attracting adjacent ions, form their own trap must be rejected. Quantitative analysis by Markham and Seitz (1948) for alkali halides shows that the traps are not deep enough; their calculations may be extended to the case of aluminium oxide.

In the absence of an external field there is an equilibrium concentration of interstitial Al^{3+} ions present in the film. Owing to the open nature of the oxide film, the application of an external field will disrupt this equilibrium and aluminium ions will be more concentrated in the direction of the negative field (i.e. near the electrolyte). It has not proved possible to demonstrate this change in equilibrium conditions, although the inertia effects found in the electron current may possibly be related to it.

Two other properties of the insulating oxide film which may be related to the presence of trapped charges have not been discussed in this paper: the rectification properties of the film and its reforming characteristics on standing.

ACKNOWLEDGMENTS

The author wishes to thank the Director of the Atomic Energy Research Establishment for permission to publish this paper; appreciation is also due to Messrs. A. H. Tilsley and L. M. Neal for assistance with the experimental details.

REFERENCES

CHARLESBY, A., 1953, *Proc. Phys. Soc.* B, **66,** 317.

GÜNTHERSCHULZE, A., and BETZ, H., 1931, *Z. Phys.*, **68,** 145; 1932, *Ibid.*, **73,** 586; 1934 a, *Ibid.*, **91,** 70; 1934 b, *Ibid.*, **92,** 467.

HERMANN, W., 1940, *Wiss. Veröff. Siemens-Werk.*, 188.

MARKHAM, J. J., and SEITZ, F., 1948, *Phys. Rev.*, **74,** 1014.

Electrode Contamination in Electron Optical Systems

By K. M. POOLE*
The Clarendon Laboratory, Oxford

MS. received 17*th December* 1952, *and in amended form* 23*rd March* 1953

Abstract. The results of a brief study of the origin and mode of formation of high resistance films on electrodes of demountable systems are presented. Evidence is given that the films are produced by the dissociation of organic molecules into radicals by electron bombardment and the polymerization of the latter. It is shown that, except where the partial pressure of organic molecules is 10^{-4} mm Hg or more, the process takes place in condensed layers of organic substances, that the nature of the electrode material is not important, and that a beam potential of a few volts is sufficient. Above 300 v beam potential only a low resistance deposit is formed; no deposit is found if a temperature of 250°c or more is maintained. A qualitative comparison of the deposits originating from various vacuum seals in common use is described.

§ 1. INTRODUCTION

In a recent attempt to obtain a beam of electrons homogeneous in energy to within narrow limits in a demountable vacuum system, the appearance of high resistance films on various electrodes led to instability in the system. The appearance of black deposits and of films showing interference colours in electron optical systems is well known, but there seems to have been little systematic investigation of the origin and formation of the deposits. Stewart (1934) measured the breakdown temperatures, breakdown voltage and resistance of a typical film.

In a gas discharge, organic molecules may be ionized, the ions may lose charge to form free radicals, and the radicals either may be decomposed further to give 'molecules' consisting of a carbon chain with occasional hydrogen atoms attached or may form stable long chain molecules by polymerization (Linder and Davies 1931, Conrad 1934, Harkins 1934). König and Helwig (1951) showed that a solid film produced by polymerization of benzene vapour in a gas discharge had a conductivity of $10^{-9}\,\Omega^{-1}\,\mathrm{cm}^{-1}$, which was increased by heating in vacuum to 600°c to $10\,\Omega^{-1}\,\mathrm{cm}^{-1}$.

§ 2. THE FORMATION OF DEPOSITS

The investigation was confined to systems evacuated by a mercury diffusion pump (for contamination from oil diffusion pumps see Webster *et al.* 1932, Hillier 1948). In all the experiments to be described a liquid oxygen trap was used to prevent the access of mercury vapour to the experimental chamber: the demountable joints under test either were a part of this chamber or were between it and the trap.

---

* Now at Bell Telephone Laboratories, Murray Hill, New Jersey, U.S.A.

(i) *Measurement of Breakdown Temperatures*

A number of experiments was designed to measure the breakdown temperature of the films produced by a low energy electron beam in a glass vacuum system using greased joints. The form finally used consists of a target mounted with a soft iron counterpoise on a pivoted arm, so that it could be moved by an external magnet from one position in front of an electron gun to a measuring position. At this position a copper electrode could be moved from outside to press lightly on the surface over an area of approximately 9 mm^2. The resistance between this electrode and the target, due to any surface film formed, was measured, a potential of 1 volt or less being applied. The failure of an earlier form in which silver was deposited on to a restricted area of the film to facilitate the measurement of the specific resistance of the material was presumably due to the presence of holes in the film leading to local short circuiting. In the final form, only an approximate value can be found for the specific resistance. The target electrode can be heated at any position by the use of the radiation heating arrangement described by Leemans and Kompfner (1951) and its temperature estimated from the power input to the lamp, a subsequent calibration using a similar plate suitably mounted being made to find the relation between lamp power and target temperature. Values found were

Beam of approximately 2 mA at 100 v for 20 hours

Conductivity of deposit	10^{-9} to 10^{-10} Ω^{-1} cm^{-1}
Breakdown temperature in vacuum	600 ± 50°c
Breakdown temperature in air	450 ± 50°c

These values are consistent with those of König and Helwig and point to a possible identity between the deposits in the present case with those in a gas discharge in organic vapours. This assumption of identity was further tested by a search for free radicals in a system containing an electron beam and the vapour of vacuum grease.

(ii) *Presence of Organic Radicals*

The apparatus for this test consisted of a connected pair of similar tubes. A lead mirror had been formed in each by evaporating lead tetra-ethyl on to a warmed surface, and each contained the vapour produced by heating Apiezon 'L' grease in a side arm. The equivalent nitrogen pressure in the system as measured on an ionization gauge before heating the side arm was approximately 10^{-6} mm Hg; on warming to approximately 80°c a pressure of 2×10^{-4} mm Hg was reached and maintained for several days. In one of the tubes an electron beam was operated at a voltage of 1 kv, the current being approximately 1 mA. After 48 hours operation the appearance of the lead mirror in the tube containing the beam had changed, the reflectivity having decreased and the transparency increased compared with that in the other tube. On warming both films to between 150 and 200°c under a pressure of 4×10^{-6} mm the comparison mirror was not affected while that in the live tube evaporated. Thus this film was more volatile than the original lead, which is true of the organic, but not the inorganic lead compounds. It is deduced that free radicals are present in such a system.

(iii) *Distribution of Deposits*

If it is assumed that the formation and polymerization of the free radicals takes place in an original film of organic material condensed on the electrodes, then the resulting deposit would be expected to appear only on electrodes intercepting part of the electron beam. In the experiment to measure the breakdown temperature of the contamination films a more widespread deposit had been observed. Free radicals formed from organic vapour in the path of the beam would subsequently polymerize on any surface, but the pressure of some 10^{-4} mm Hg required before this effect is significant is far higher than that in any practical electron optical vacuum envelope. This discrepancy was studied in the apparatus shown in fig. 1.

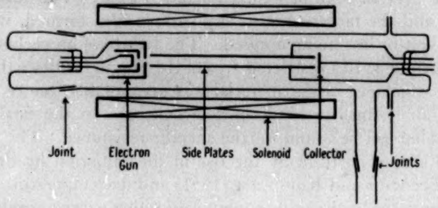


Fig. 1.

An electron beam at 750 v was directed between a pair of parallel plates, originally at anode potential, on to a collector. The beam was constrained by an axial magnetic field extending the whole length of the system so that no part of it could be intercepted by the plates. Two greased joints were used in the vacuum system. With a beam current of 1·2 mA running for 20 hours, a deposit showing up to three spectral bands of interference colours was produced on the side plates, the maximum thickness occurring near the gun aperture and near the collector. Away from these points the deposit was very slight, confirming that at normal operating pressures the number of free radicals formed in the vapour phase is negligibly small. The experiment was repeated with the collector still at anode potential, but the side plates 20 v below this. The beam was run for 30 hours at an average current of 1·4 mA; on examination it was found that there was no heavy deposit on any part of the side plates. It seems that secondary electrons striking an electrode with their energy of emission, that is with energies of the order of 5 electron volts, can form high resistance deposits; this suggests that free radicals can be formed directly with no intermediate ionization stage. This hypothesis is sufficient to account for all the cases of a widespread distribution of deposits observed in the earlier experiments, and receives some corroboration from the possibility of the contamination of a mercury discharge tube by inorganic vapours where the discharge is spectroscopically pure, as observed by Dr. P. C. Thonemann (unpublished).

§ 3. SOURCES OF CONTAMINATION

It appears, then, that the mechanism of formation of the high resistance films is the production of free radicals in a layer of organic material on an electrode, followed by the polymerization of these radicals to form stable compounds of high resistance, an electron energy of a few volts being sufficient. If this is so, any organic material used as a seal is likely to give contamination unless its vapour pressure is much lower than that of the Apiezon grease used in the experiments described above, and the nature of the electrode material is not likely to be a significant variable.

(i) *Method of Study*

To compare the contamination from different sealing materials and determine any upper limit to the range of beam energies in which their production is possible, the apparatus shown in fig. 2 was used. The target plate is bombarded by electrons from the tungsten filament which have passed through the small aperture in the anode; the vacuum seal is made with the material under test. The target and anode potentials are derived from a low resistance divider, the target potential being half the anode potential. The test depends on the fact that the electrode film is always found to have a secondary emission factor passing through unity at an electron energy between 300 and 600 ev. A film having been formed on the target by electron bombardment, its presence is demonstrated by the appearance of hysteresis in the relation between target current and potential, the normal form of relation being as shown in fig. 3. On increasing the potential from zero the potential of the front surface of the film remains lower than the applied potential by an amount depending on the film resistance. This continues until the potential of the front surface approaches that giving unity secondary emission factor, when it rises rapidly to a new stable value above the applied target potential, the target current becoming negative and the secondary electrons being collected by the anode. If the applied potential is now reduced, the current remains negative until the surface potential again approaches that giving unity secondary emission factor, the applied potential being well below this, when the current again changes sign very rapidly. The separation of the two branches of the curve gives an indication of the resistance of the film; a series of films, formed from a sealing material at various beam voltages, can be tested to check the possible existence of an upper limiting value above which the film is destroyed by bombardment as it is formed. The apparatus was set up with a mercury pump and liquid air trap, the only joint between the electrodes and the trap being that under test; the electrode structure was cleaned between successive runs

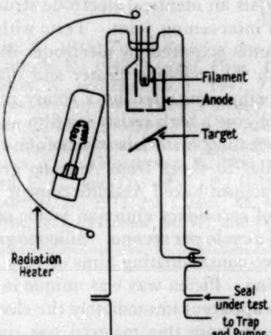


Fig. 2.

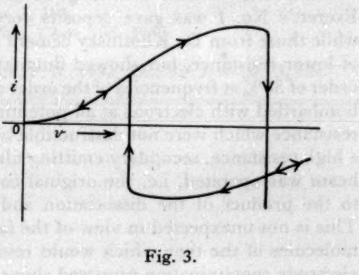


Fig. 3.

with one seal by baking with the radiation heater, and both the electrodes and the liquid air trap were replaced before starting tests on a new seal material, the new electrodes being vacuum baked before use.

(ii) *Results*

The first series of tests was with Apiezon grease as the sealing material and targets of copper, nickel, Ferry (non-ferromagnetic copper–nickel alloy), silver and stainless steel. No variation with target material was found; a Ferry target was used in subsequent tests. It was then established that films can be formed from Apiezon grease condensates at all electron beam energies up to approximately 300 v. At higher potentials a black deposit is obtained without any high resistance film having been formed. It was also established that contamination can be prevented by maintaining a target temperature of at least 250°c during, and for a few minutes before, the operation of the electron beam. The radiation heater was used as the source of heat in this test.

Before testing other vacuum sealing materials an identical electrode structure was sealed onto the pumping system with no intervening joint. Tests with this arrangement showed that grease from the hands acquired by electrodes during assembly could be eliminated by baking with the radiation heater and that oil vapour which enters the system while it is being evacuated by a rotary pump, i.e. before the diffusion pump is in use, did not give a high resistance film.

In a series of tests on a range of vacuum sealing materials it was found that Everett's No. 1 wax gave deposits very similar to those from Apiezon grease, while those from De Khotinsky cement and vacuum baked Araldite cement were of lower resistance, but showed fluctuations of secondary emission factor of the order of 30% at frequencies of the order of half a cycle per second. Silicone grease bombarded with electrons at all potentials gave contaminating films of very high resistance which were not destructible by baking. Picien wax was unique in that a high resistance, secondary emitting film was apparent immediately the electron beam was operated, i.e. the original condensate from this material was similar to the product of the dissociation and polymerization process in other cases. This is not unexpected in view of the fact that Picien wax consists of long chain molecules of the type which would result from the mechanism of formation of electrode contamination proposed above.

The only seal of those tested which might be satisfactory at beam potentials below 300 v was a neoprene gasket joint. This showed no high resistance film under the conditions of these tests, though some deposit is formed, characterized by a high and fluctuating secondary emission factor and the appearance of a black layer on the target. This type of contamination is only important in electron optical systems in which the fluctuations in electrode current lead to prohibitive voltage fluctuations. The particular example which was the occasion for this study was one such system.

§ 4. CONCLUSION

It is concluded that high resistance films in demountable electron optical apparatus are formed from organic condensates on the electrodes by a process of dissociation of organic molecules into free radicals under electron bombardment, followed by polymerization to give stable long chain molecules. At beam potentials above about 300 v the film is broken down as fast as it is formed, leaving only a black deposit—probably chains of carbon atoms with occasional hydrogen atoms attached—which, while normally being of no importance to the optics of a system, may be serious as a source of specimen contamination in electron microscopes (see Hillier 1948, Davidson and Hillier 1947, Cosslett 1947). At lower potentials most common vacuum seals give high resistance films, the only exception found being a neoprene gasket joint used with no grease. Contamination may also be avoided by heating the whole electrode system to at least 250°c.

ACKNOWLEDGMENTS

The author wishes to thank Professor Lord Cherwell for the facilities of the Clarendon Laboratory, Dr. R. Kompfner and Dr. P. Shaw for their interest in the work, and Dr. G. Liebmann and his co-workers at Associated Electrical Industries, Aldermaston, for a helpful discussion on the origin of deposits appearing outside the direct line of a beam.

REFERENCES

CONRAD, R., 1934, *Trans. Faraday Soc.*, **30**, 215.
COSSLETT, V. E., 1947, *J. Appl. Phys.*, **18**, 844.
DAVIDSON, N., and HILLIER, J., 1947, *J. Appl. Phys.*, **18**, 499.
HARKINS, W. D., 1934, *Trans. Faraday Soc.*, **30**, 221.
HILLIER, J., 1948, *J. Appl. Phys.*, **19**, 226.
KÖNIG, H., and HELWIG, G., 1951, *Z. Phys.*, **129**, 491.
LEEMANS, A., and KOMPFNER, R., 1951, *Vacuum*, **1**, 203.
LINDER, E. G., and DAVIES, A. P., 1931, *J. Phys. Chem.*, **35**, 3649.
STEWART, R. L., 1934, *Phys. Rev.*, **45**, 488.
WEBSTER, D. L., HANSEN, W. W., and DUVENECK, F. B., 1932, *Rev. Sci. Instrum.*, **3**, 729.

Additivity of Colour Equations

By P. W. TREZONA*
Technical Optics Section, Imperial College, London

MS. received 25*th February* 1953

Abstract. A colour can be matched by an additive mixture of three coloured lights, the most suitable being specified red, green and blue stimuli. If c_1 units of one colour are matched by r_1, g_1 and b_1 units of red, green and blue respectively, and c_2 units of another colour by r_2, g_2 and b_2 units, then according to the law of additivity of colour equations the combined colour may be matched by r_1+r_2, g_1+g_2 and b_1+b_2 units respectively. It is the purpose of this work to investigate this law.

Additivity of colour equations is implicit in all colorimetric calculations. The law is based on the results of early work done with the comparatively crude apparatus of the time, but recently these results have been questioned.

The experimental work was performed on the Wright colorimeter, where colour matches were made on pairs of stimuli separately and in combination. The results showed some deviations from additivity, but on the whole these were small. Large deviations appear to be associated with poor discrimination.

§ 1. Introduction

The law of additivity of colour equations can be stated as follows : if c_1 units of a colour C_1 are matched by r_1, g_1 and b_1 units of the three matching stimuli R, G and B respectively and if c_2 units of another colour C_2 are matched by r_2, g_2 and b_2 units of the same three matching stimuli, then an additive mixture of c_1 units of C_1 and c_2 units of C_2 is matched by r_1+r_2, g_1+g_2 and b_1+b_2 units of the matching stimuli. Algebraically,

$$\begin{aligned}&\text{if} && c_1C_1 = r_1R + g_1G + b_1B\\ &\text{and} && c_2C_2 = r_2R + g_2G + b_2B\\ &\text{then} && c_1C_1 + c_2C_2 = (r_1+r_2)R + (g_1+g_2)G + (b_1+b_2)B.\end{aligned}$$

Another form of this law is the process of finding the mixture of two colours represented by points on a colour triangle by the centre of gravity rule. Newton (1704) first suggested this process, although his use of a colour circle invalidated its application. The introduction of the colour triangle by Young (1807) allowed Helmholtz (1911) to apply the centre of gravity rule to it for points both inside and outside. Helmholtz accompanied this by a mathematical treatment using Grassman's third law to justify adding together different amounts of the same colour. This step is invalidated by the fact that Grassman's third law (1853) itself relied on the ability to add different amounts of the same colour; Helmholtz's work may, instead, be taken as a demonstration of the equivalence of the algebraic and the ' centre of gravity ' forms of additivity.

---

* Now at Queen Elizabeth College, Campden Hill Road, London W.8.

On the experimental side, Helmholtz's contemporary, Maxwell (1890), showed the validity of the centre of gravity rule using rotating discs with coloured sectors. A certain colour was first matched directly by three matching stimuli. Then it was matched by three other stimuli and transformed, using this rule, to the first set. He found that "theory agrees with calculation always within 0·012 of the whole and sometimes within 0·002". König (1903), using spectral lines, tested the additivity of similar colours or, what is the same thing, the maintenance of a colour match with change of intensity. His matches held over a considerable range.

On the basis of this work additivity was generally regarded as holding under ordinary conditions. Recently, however, Blottiau (1947) has claimed to have found large deviations. He used a Donaldson colorimeter with a 2° field. In order to make a match he desaturated a spectral colour by one of the matching stimuli R, G or B or by a white W, as convenient. The same amount of the spectral colour was added in turn to three different amounts of the desaturating colour; fig. 1 shows how the apparent position of the spectral colour alters for one of his observers. Although the spectral colour nearly always appears to go further from the triangle as the amount of desaturating colour increases, other observers show a movement in the opposite direction. Still others showed movement first in one direction and then in the other. By taking the mean of all observers he found no appreciable change. For the combined colour, inside the triangle, a considerable difference is seen between its position by calculation and that by direct matching. But the biggest discrepancy is not in the red, the colour of the varied matching stimulus, but in the unvaried blue.

The present investigation was undertaken in an attempt to discover to what extent Blottiau's contradiction of the results of the early workers is valid.

§ 2. METHOD

(i) *Apparatus*

The apparatus used was the Wright colorimeter (1946) which provided a bipartite field subtending an angle of 1° 20′ at the eye. The upper half gave a mixture of the three matching stimuli R ($0{\cdot}65\,\mu$), G ($0{\cdot}53\,\mu$) and B ($0{\cdot}46\,\mu$), and their amounts could be altered by three neutral wedges adjusted by the observer. The lower half could be of any spectral colour set to a convenient intensity by a neutral filter. Also in the lower half R, G or B could be added as required for desaturating purposes; this will be called the desaturating stimulus (D.S.) The test colour in the lower half was chosen and then a colour match was made by adjustment of the matching stimuli in the upper half of the field.

(ii) *Observers*

The six observers were J.E.G., I.G.H.I., E.G.T., L.C.T. and W.D.W., who have normal colour vision, and P.W.T., who is protanomalous. E.G.T. and P.W.T. are women. I.G.H.I. and E.G.T. are Egyptians, and the other observers are British. Egyptians have spectral luminous efficiency (V_λ) curves similar to the British in the red part of the spectrum but significantly lower in the blue, the difference being attributed to a denser layer of macular pigmentation (Ishak 1952). Accompanying it is an insensitivity to varying amounts of the blue matching stimulus in a colour match. Also, in the case of I.G.H.I., a match acceptable on one occasion would not be so on another.

P.W.T. gives a protanomalous yellow match requiring too much red, but in this respect her difference from the normal is small. Her V_λ curve is very low in the red and her hue discrimination curve, although markedly different from the normal, does not correspond to an extreme case of protanomaly. In colour matching her G and B matching stimuli showed no more observational spread than those of a normal observer, but R gave much more variation.

(iii) *Observing Technique*

When the observer arrived at the colorimeter he was given a short time of dark adaptation to allow any after-images to disperse. He was then presented with a colour which he was encouraged to match within two minutes. He was allowed either to scan the field or to fixate on the centre according to personal preference. He used the R, G and B controls in turn, obtaining a match by successive approximations. Each matching stimulus was set by the 'bracketing' technique or something closely approaching it. When he was satisfied with a match, the scales were read by the operator and then altered slightly. Each match was repeated twice. A single test on additivity was made in each observing session. This usually lasted 1–1½ hours, during which time the observer was allowed a short rest.

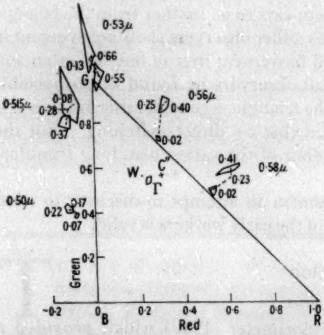


Fig. 1.

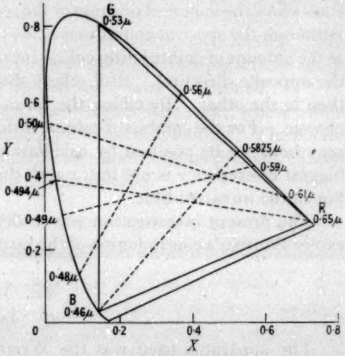


Fig. 2.

Fig. 1. Blottiau's results. For each point the quantity $(R_1+G_1+B_1)/(R+G+B)$ is given, where (R_1, G_1, B_1) are the coefficients of the desaturating stimulus and (R, G, B) those of the mixture. The convex polygons show the limits of the ten matches made. Γ gives the position of the mixture of 0·515 μ and a large amount of the desaturating stimulus R. C″ is the new position of Γ calculated from 0·515 μ combined with the least amount of R.

Fig. 2. Positions on the C.I.E. Chromaticity Chart of the colours combined for testing additivity. R, G and B are the matching stimuli.

(iv) *The Investigation*

The main investigation fell into two types. In type A two spectral colours were matched separately and in combination, neither colour being that of a colorimeter matching stimulus. In type B a spectral colour was matched with four different amounts of a desaturating stimulus (D.S.) of the same wavelength as one of the matching stimuli; these matches, together with intensity matches on the four amounts of D.S., gave a way of testing additivity. In type A a small amount of D.S. had to be added to each spectral colour in order to obtain a colour match. This amount was kept fairly low in an attempt to minimize its effects and make type A as different as possible from type B.

The actual colour combinations chosen are shown joined by dotted lines in fig. 2. Their choice was largely governed by an attempt to cover the colour domain as evenly as possible on a just-discriminable step basis. The use of complementary spectral radiations for the colours to be combined meant that they differed from each other as much as possible. In type A their relative amounts were chosen to make the combination a fairly white colour.

The luminance level was one of convenience. It varied with the colour matched but was usually between 30 and 300 cd/m^2. The corresponding retinal illumination values, which allowed for the area of the exit-pupil of the colorimeter, were between 100 and 1000 photons.

§ 3. Results and Discussion

(i) *Treatment of Results*

From the results it must be decided whether a comparison between two quantities shows a significant difference or whether this difference must be ascribed to experimental error; one of these quantities is a sum derived from two separate matches and the other is a single value obtained when the colours are combined. For every test of additivity there will be three such 'sum' and 'mixture' values to compare—one for each of the red, green and blue matching stimuli. Weber's law justifies this comparison taking the form of a percentage deviation. If the intensity is I and a just-discriminable intensity change is ΔI, then the law $\Delta I/I=$ constant becomes $\sigma/I=$ constant where σ represents the experimental error. This will be true as long as instrumental error is small (as it is in this case) so that $\sigma \propto \Delta I$. Experimental error expressed in this form will be independent of the intensity. It follows that a quantity which has to be classed as either experimental error or not must be expressed in this same form. Prompted by this idea (but also bearing in mind that with three stimuli Weber's law may not hold for each) the percentage deviation was defined for R as

$$\frac{(r'+r'')-r}{\frac{1}{2}(r'+r''+r)} \times 100.$$

Similar expressions are used for the green and blue matching stimuli.

For type A and the test on, say, $0{\cdot}49\,\mu + 0{\cdot}59\,\mu$, r is the amount of red in the mixture, r' the amount in $0{\cdot}59\,\mu$ and r'' the amount in $0{\cdot}49\,\mu$. The last is obtained by using a D.S. and will be negative. For type B and the test on, say, $0{\cdot}494\,\mu + 0{\cdot}65\,\mu$, r is the amount of red in the colour match, r' the amount in the intensity match on the D.S. ($0{\cdot}65\,\mu$) alone and r'' the derived (negative) amount of red in $0{\cdot}494\,\mu$. The small negative quantity r'' was obtained from the colour and intensity matches when the smallest amount of D.S. was used. For this colour combination of $0{\cdot}494\,\mu + 0{\cdot}65\,\mu$ four colour and four intensity matches were made, thus giving three percentage deviations all for the red matching stimulus. Since only the red stimulus in the test combination was changed, no variation would be expected in the green and blue matching stimuli; this, however, has to be checked. In a similar way, three percentage deviations were found for the blue matching stimulus in $0{\cdot}5825\,\mu + 0{\cdot}46\,\mu$.

An alternative, but less satisfactory, way of dealing with type B results is to give four different estimations of the amount of negative red in $0{\cdot}494\mu$, corresponding to the four amounts of R. This negative red is a difference between two quantities; when the amount of the D.S. is small its determination will be accurate, but for large D.S's two large quantities are subtracted to give a small difference, and experimental error will completely swamp the observations. As an example of how experimental error increases with increase in the amount of D.S. (a fact consistent with Weber's law) the results of one occasion are given. The standard deviation σ varied, as the amount of D.S. was increased, from 0·8 to 1·7 to 8·5.

(ii) *Results expressed as Percentage Deviations and Statistically*

Percentage deviations of type A are shown in table 1, where every individual test is given separately. It will be seen that deviations are never greater than 10% in the red, 9% in the green and 30% in the blue; usually they are a great deal less. The arithmetic means of all the percentage deviations are: R, −1%; G, 0%; B, −3%.

Table 1. Percentage Deviations from Additivity—Type A

		$0{\cdot}49\mu+0{\cdot}59\mu$			$0{\cdot}48\mu+0{\cdot}56\mu$			$0{\cdot}50\mu+0{\cdot}61\mu$		
		R	G	B	R	G	B	R	G	B
Normal British	J.E.G.	−3	−2	−5	−10	−9	+6	−7	+1	−4
		−4	+3	−3	+3	−2	−7	+1	+1	−5
	L.C.T.	−4	+6	−3	−9	−1	−8	+4	+3	−17
		−1	+3	−6	−6	+3	+30	−6	−3	−7
		−4	−1	+7						
		−5	−1	−4						
	W.D.W.	+5	+1	−9	−4	+2	−12	+1	−4	−6
		−1	−1	−7	−1	−1	0	0	−1	−6
		0	+1	−5	0	+1	+12	+1	−5	−12
		+1	−1	−4	−8	+1	−6			
Normal Egyptian	I.G.H.I.	+1	+5	+2	−6	+2	−520	−5	−5	−19
		−1	−2	+8	−1	+5	−450	+2	+3	−30
		+5	+3	+8						
		−6	−3	−18						
	E.G.T.	−5	−1	+2	+1	−1	−87	+10	+4	+23
		0	0	+7	−3	−2	+24	+8	0	+13
		+4	+2	−8	−6	−1	−14	+1	−2	+8
Protanomalous British	P.W.T.	+8	0	0	+20	−5	−3	+12	+1	−7
		+11	−1	−9	+13	+1	−7	+57	−5	−2
		+26	−1	−8	−2	+2	+1	+27	−17	+4
		+21	−2	−11				+15	−7	+13

Table 2 shows deviations of type B for the varied matching stimulus. The unvaried matching stimuli showed no change. R_3 refers to the case when the greatest amount of D.S. was used and R_1 to the smallest but one, the smallest having been used to determine the small amount of negative red in $0{\cdot}494\mu$. For $0{\cdot}494\mu+0{\cdot}65\mu$ the red deviations are never greater than 13%. The means are: R_1, +3%; R_2, +2%; R_3, +3%. For $0{\cdot}5825\mu+0{\cdot}46\mu$ the largest deviation is

33%. The mean blue deviations are: B_1, −3%; B_2, +6%; B_3, +11%. In the above figures, all results of P.W.T., who is protanomalous, were ignored. Results for I.G.H.I. and E.G.T. were ignored for the blue matching stimulus only.

Table 2. Percentage Deviations from Addivitity—Type B

The Varied Matching Stimuli

		$0{\cdot}494\,\mu+0{\cdot}65\,\mu$			$0{\cdot}5825\,\mu+0{\cdot}46\,\mu$		
		R_1	**R_2**	**R_3**	**B_1**	**B_2**	**B_3**
Normal British	J.E.G.	+7	+2	+12	−7	+15	+27
		0	−2	−1	−2	+12	+33
					−7	+16	+19
	L.C.T.	−4	−1	−6		−8	−3
		−2	+3	−6	−15	−5	−1
		+13	+8	+1			
	W.D.W.	+6	+4	+6	−1	+9	+2
		+6	+1	+6	+2	0	+3
		+1	+3	+1	+6	+9	+6
Normal Egyptian	I.G.H.I.	+7	+5	+3	+27	+47	+45
		−1	+4	+10	+8	+31	+49
		−1	+2	+1			
	E.G.T.	+5	+4	+4		+35	−1
		−5	−1		−27	+17	+17
		+8	0		+17	+26	+30
					+8	+16	+31
Protanomalous British	P.W.T.	−1	−15	−11	+7	+6	+13
			+8	−8	+3	+1	+12
		+11	−3	−5	+3	−6	0

In order to judge each deviation relative to its experimental error a statistical analysis was undertaken. Table 3 gives results of Student's *t*-test when all results for one observer on one pair of colours are combined. A tick indicates the passing and a cross the failure of the test at the 1% significance level. A positive sign is given by the cross when the failure is in such a direction that the sum is larger than the mixture; a negative sign indicates that the sum is significantly smaller than the mixture.

From the three tables it will be seen that there is a general negative tendency for type A, the sum being smaller than the mixture. This occurs for the blue matching stimulus, to a lesser extent for the red and not at all for the green. In type B on the whole there is a general tendency to make the sum larger than the mixture. Significant deviations appear to be associated in part with poor discrimination. One example of this is P.W.T's test on $0{\cdot}49\,\mu+0{\cdot}59\,\mu$, which was significant even at the 0·01% level.

(iii) *Results expressed on the Chromaticity Chart*

Plotting results on the chromaticity chart is unsatisfactory from some points of view since it does not give all the available information. For instance, it is possible for all three matching stimuli to have large positive deviations and for it not to show at all on the chromaticity chart. However, the method has the advantages of being simple and clear and fig. 3 shows the positions of the mixture M and the sum S for three sets of observations by W.D.W.

Table 3. Significance Test 1% Level

	$0{\cdot}49\,\mu+0{\cdot}59\,\mu$			$0{\cdot}48\,\mu+0{\cdot}56\,\mu$			$0{\cdot}50\,\mu+0{\cdot}61\,\mu$		
	R	**G**	**B**	**R**	**G**	**B**	**R**	**G**	**B**
J.E.G.	✓	✓	✓	✓	$\times^{-}$	✓	✓	✓	✓
L.C.T.	$\times^{-}$	✓	✓	✓	✓	✓	✓	✓	✓
W.D.W.	$\times^{+}$	✓	$\times^{-}$	$\times^{-}$	✓	✓	✓	✓	$\times^{-}$
I.G.H.I.	✓	✓	✓	✓	$\times^{+}$	$\times^{-}$	✓	✓	$\times^{-}$
E.G.T.	✓	✓	✓	✓	✓	✓	$\times^{+}$	✓	$\times^{+}$
P.W.T.	$\times^{+}$	✓	$\times^{-}$	$\times^{+}$	✓	✓	$\times^{+}$	$\times^{-}$	✓
Five normal observers combined	✓	✓	$\times^{-}$	$\times^{-}$	✓	$\times^{-}$	✓	✓	$\times^{-}$

	$0{\cdot}494\,\mu+0{\cdot}65\,\mu$			$0{\cdot}5825\,\mu+0{\cdot}46\,\mu$		
	R_1	**R_2**	**R_3**	**B_1**	**B_2**	**B_3**
J.E.G.	✓	✓	✓	✓	$\times^{+}$	$\times^{+}$
L.C.T.	✓	✓	✓	$\times^{-}$	✓	✓
W.D.W.	$\times^{+}$	$\times^{+}$	$\times^{+}$	✓	$\times^{+}$	$\times^{+}$
I.G.H.I.	✓	$\times^{+}$	✓	✓	$\times^{+}$	$\times^{+}$
E.G.T.	$\times^{+}$	✓	✓	$\times^{+}$	$\times^{+}$	$\times^{+}$
P.W.T.	✓	✓	$\times^{-}$	✓	✓	$\times^{+}$
Five normal observers combined	$\times^{+}$	$\times^{+}$	$\times^{+}$	$\times^{+}$	$\times^{+}$	$\times^{+}$

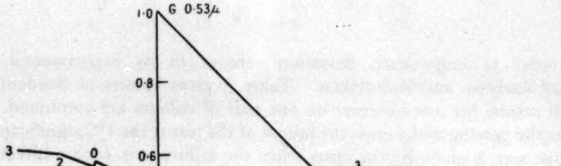


Fig. 3. The Wright Chromaticity Chart for observer W.D.W. showing the positions of the mixture M and sum S and the apparent changes in position of $0{\cdot}494\,\mu$ and $0{\cdot}5825\,\mu$. 0 corresponds to the least amount of D.S. and 3 to the greatest.

A consideration of figs. 1 and 3 will give a comparison of the present work with that of Blottiau. Both works show large changes in the apparent position of spectral colours, but, as was pointed out above, this is much less significant than it appears at first sight because of the large experimental error. Blottiau's position of C″ shows large deviations in the unvaried matching stimuli; this is not in keeping with the results of the present investigation. One comparison alone is rather unsatisfactory, but unfortunately Blottiau published no more results of this type.

Figure 4 shows the mean mixture–sum lines for the normal observers on the C.I.E. chromaticity chart. In plotting these, the 'mixture' positions are derived from W.D.W.'s results, and the 'sum' points are placed to give the appropriate deviation. This figure enables the M–S lines to be compared with lines representing three times just-discriminable colour differences at constant luminance (Wright 1944). On the whole deviations are small compared with the just-discriminable colour differences.

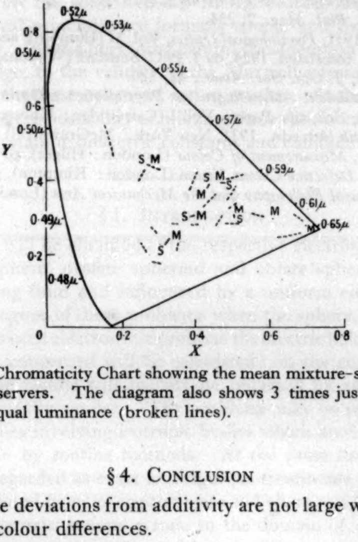


Fig. 4. The C.I.E. Chromaticity Chart showing the mean mixture–sum (M–S) lines for all the normal observers. The diagram also shows 3 times just-discriminable colour difference at equal luminance (broken lines).

§ 4. CONCLUSION

(i) On the whole deviations from additivity are not large when compared with just discriminable colour differences.

(ii) The average values found were :

Type A			Type B					
R	G	B	R_1	R_2	R_3	B_1	B_2	B_3
−1%	0%	−3%	+3%	+2%	+3%	−3%	+6%	+11%

(iii) The results were roughly in agreement with those of Blottiau, but his method of presentation exaggerated the deviations.

(iv) In type A there is a general negative tendency, the sum being smaller than the mixture. This occurs for the blue matching stimulus, to a lesser extent for the red and not at all for the green.

(v) On the whole there is a positive tendency for type B, the sum being larger than the mixture. The unvaried matching stimuli show no change with increasing amounts of D.S.

(vi) Large deviations from additivity are associated with poor discrimination.

ACKNOWLEDGMENTS

The author is very grateful to Professor W. D. Wright for suggesting the problem, for many helpful suggestions and criticisms and for acting as an observer. Thanks are also due to Dr. L. C. Thomson for his helpful suggestions and for acting as an observer, and to Mrs. E. G. Wassef, Mr. J. E. Gibbs, Dr. I. G. H. Ishak and Mr. W. S. Yang for observing and recording readings. The author gratefully acknowledges the financial support of the Medical Research Council.

REFERENCES

BLOTTIAU, F., 1947, *Rev. Opt.* (*Théor. Instrum.*), **26,** 193.

GRASSMANN, H., 1853, *Phil. Mag.*, **7,** 254.

VON HELMHOLTZ, H., 1911, *Physiological Optics*, Vol. II (Hamburg and Leipzig : Verlag von Leopold Voss), translated, 1924, by J. P. C. Southall (Wisconsin : Opt. Soc. Amer.).

ISHAK, I. G. H., 1952, *J. Opt. Soc. Amer.*, **42,** 529.

KÖNIG, A., 1903, *Gesammelte Abhandlugen zur Physiologischen Optik* (Leipzig : Barth).

MAXWELL, J. C., 1890, *Scientific Papers*, Vol. I (Cambridge : University Press).

NEWTON, I., 1704, *Optik* (4th edn., 1931, New York : McGraw-Hill).

WRIGHT, W. D., 1944, *Measurement of Colour* (London : Hilger), p. 184; 1946, *Researches on Normal and Defective Colour Vision* (London : Kimpton), p. 45.

YOUNG, T., 1807, *Natural Philosophy and the Mechanical Arts* (London : Joseph Johnson), p. 440.

Some Electrostatic and Steady-Current Problems Involving Anisotropic Bodies*

By R. CADE

University College of the West Indies, Jamaica

Communicated by R. Fürth; *MS. received* 19*th December* 1952

Abstract. Solutions are given of six new electrostatic and steady-current problems involving anisotropic spheres and spheroids influenced by uniform electric fields. In the electrostatic problems derivations are given of the potentials and the mechanical action on the solid bodies; in the corresponding steady-current problems the physical principles are formally similar and the results are usually obtained by analogy. At the beginning careful consideration is given, in the light of modern knowledge, to the validity of the Maxwell stresses, which are subsequently used to obtain the mechanical action results. At the end a brief critical discussion is made of the possibility of utilizing these results experimentally for the determination of crystalline dielectric constants and conductivities.

§ 1. Introduction

The solutions will be obtained of the respective electrostatic problems of an anisotropic sphere, prolate spheroid and oblate spheroid immersed in a non-conducting fluid and influenced by a uniform electric field, and the steady-current analogues of these problems when the sphere, spheroids and fluid are conducting. In each electrostatic problem the electric field and the mechanical action on the body concerned will be calculated; in the corresponding steady-current problem the results will usually be obtained by analogy, the physical principles being formally similar. The problems may be regarded as generalizations of similar ones involving isotropic bodies which are either well known or are readily tractable by routine methods. At the same time, the electrostatic problems may be regarded as exact and rigorous treatments of special cases of a general problem treated from an approximate and elementary standpoint, and, as will be seen, with certain serious errors, in the domain of crystal physics. In this field the steady-current problems, although of importance, do not appear to have been treated at all.

The author has solved the problems in detail elsewhere (Cade 1952 b). For conciseness there will be a minimum description of mathematical steps here, and detailed discussion will be omitted except where it is of special significance. Special consideration is given to the possibility of experimental application of the results since they might prove to be of some importance to crystal physics.

One of the difficulties which occurred in the course of the work cited was the question of the theory of mechanical action in dielectrics. The subject was then highly controversial, and still is to some extent, although the author believes that the essential difficulties have been overcome by recent work. The earliest theory was that of the Maxwell stresses, which was intuitive rather than deductive. Deductive theories have followed two definite lines. The Helmholtz theory, which is the 'conventional' one in the sense that it is usually quoted in the treatises (e.g. Abraham and Becker 1932, Jeans 1943, Smythe 1939), presupposes a theory of energetics and derives the electric force on a dielectric by a variational method. The other point of view, propounded by Livens (1926) and developed by Smith-White (1949) and Brown (1951), is to obtain the electric force by direct appeal to the Poisson–Kelvin theory of dielectric polarization and to derive the energetics *a posteriori*. The principles of this method are simple, and although the development of the general theory was largely stimulated by criticisms that were made of the Helmholtz theory, it is of interest to note that the principles were actually used much earlier in crystal physics, although with certain errors previously mentioned.

---

* This paper forms part of a thesis accepted by the University of London for the Ph.D. degree.

There can be no doubt that the Helmholtz theory is unsound. The first criticism was given by Larmor (1897), and others are to be found in the literature of Livens, Smith-White and Brown. When the problems we are concerned with were first treated (Cade 1952 b), Brown's paper (1951) had not appeared, and a fundamental error had been discovered in the Livens–Smith-White theory (Cade 1951, 1952 a). The author's belief in the essential soundness of Livens' and Smith-White's approach was expressed, but even after the correction of the error in the theory, application to the isotropic analogue of one of the problems gave results in flagrant discord with experiment (Cade 1952 b). Brown (1951), however, attacked the whole problem afresh, and not only avoided the error of Livens and Smith-White, but treated a new matter, namely the modification of the hydrostatic stresses arising from the electrical force in a fluid dielectric. That this was a possible solution to the problem was suggested by the author (Cade 1952 b) but not worked out. The result obtained by Brown for the case of a uniform, linear, incompressible, fluid dielectric is—remarkably enough after the long controversy—that the mechanical action is given by the Maxwell stresses.

In the previous work calculations of the mechanical action were made using both the Helmholtz theory and the modified Livens–Smith-White theory. The situation having been clarified by Brown, the Maxwell stresses will be used exclusively in the present work. This means that we must emphasize the sole applicability of the results to fluid dielectrics which are uniform, linear and incompressible. These conditions are necessary *a priori*, however, for the potential problems to be tractable, so that actually no new restriction upon the validity of the results is implied. Compressibility is a consideration new to the theory of mechanical action in dielectrics, but of course, unless one is considering an incompressible dielectric, the potential problem is complicated by electrostriction and has no exact solution. We may conclude that the results are valid to a good approximation for liquid dielectrics. Brown derives entirely new mechanical action formulae for compressible dielectrics, which, in practice, are gases, but for practical purposes it is usually sufficient to regard a gaseous dielectric as empty space for results such as we shall obtain. In any case, for small fields, electrostriction is so small that the dielectric may be regarded as incompressible without serious error. The limitation upon the usefulness of our results is therefore slight.

§ 2. Anisotropic Non-Conducting Sphere

We consider a uniform, linear, anisotropic dielectric with symmetric dielectric tensor

$$\epsilon_i \equiv \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ & \epsilon_{yy} & \epsilon_{yz} \\ & & \epsilon_{zz} \end{bmatrix} \qquad \ldots\ldots(1)$$

relative to the cartesian coordinate axes, and suppose that a sphere of radius a of this material is immersed in a fluid dielectric with dielectric constant ϵ_o which, before the immersion, was influenced by a uniform electric field of strength $\mathbf{F}$. We choose as origin of coordinates the centre of the sphere and, on account of symmetry, we may allow the coordinate axes to coincide with the mutually orthogonal principal axes of the dielectric tensor without incurring any mathematical complications. The dielectric tensor thus becomes diagonal :

$$\epsilon_i \equiv \begin{bmatrix} \epsilon_x & 0 & 0 \\ & \epsilon_y & 0 \\ & & \epsilon_z \end{bmatrix}, \qquad \ldots\ldots(2)$$

and $\mathbf{F}$ will be in the direction whose direction cosines are l, m, n, say. We define spherical polar coordinates by

$$\left.\begin{aligned} x &= r\cos\theta\,, \\ y &= r\sin\theta\cos\phi\,, \\ z &= r\sin\theta\sin\phi\,. \end{aligned}\right\} \qquad \ldots\ldots(3)$$

Outside the sphere the electrostatic potential must satisfy Laplace's equation,

$$\nabla^2 V_o = 0\,, \qquad \ldots\ldots(4)$$

and inside, the equation

$$\epsilon_x \frac{\partial^2 V_i}{\partial x^2} + \epsilon_y \frac{\partial^2 V_i}{\partial y^2} + \epsilon_z \frac{\partial^2 V_i}{\partial z^2} = 0 \qquad \ldots\ldots(5)$$

(Smythe 1939). In addition, the boundary conditions

$$V_o = V_i \qquad (r=a), \qquad \ldots\ldots(6)$$

$$(\epsilon_o\,\mathbf{E}_o, \mathbf{n}) = (\epsilon_i\,\mathbf{E}_i, \mathbf{n}) \qquad (r=a) \qquad \ldots\ldots(7)$$

must be satisfied, where $\mathbf{n}$ is a unit normal, outward say, at the surface of the sphere, and $\mathbf{E}_o$, $\mathbf{E}_i$ are electric intensities related to the potentials by

$$\mathbf{E}_o = -\operatorname{grad} V_o\,, \quad \mathbf{E}_i = -\operatorname{grad} V_i\,. \qquad \ldots\ldots(8)$$

The external potential is the superposition of two parts, that due to the original field $\mathbf{F}$ and that due to the replacement of a fluid sphere by a solid one with different dielectric properties, and by virtue of (4) it is expressible in spherical harmonics. Let us assume

$$\begin{aligned} V_o = &-Fr\,\{lP_1(\cos\theta) + mP_1{}^1(\cos\theta)\cos\phi + nP_1{}^1(\cos\theta)\sin\phi\} \\ &+ \frac{F}{r^2}\{\alpha_1 P_1(\cos\theta) + \alpha_2 P_1{}^1(\cos\theta)\cos\phi + \alpha_3 P_1{}^1(\cos\theta)\sin\phi\}\,, \end{aligned} \qquad \ldots\ldots(9)$$

where the α's are constants to be determined. In general it is difficult to find solutions of (5) which satisfy specified boundary conditions, but in this simple case we will assume

$$V_i = -F(\beta_1 x + \beta_2 y + \beta_3 z)\,, \qquad \ldots\ldots(10)$$

where the β's are constants to be determined. This function is also a solution of Laplace's equation, (4), and is expressible in spherical harmonics by

$$V_1 = -Fr(\beta_1 P_1 + \beta_2 P_1{}^1 \cos\phi + \beta_3 P_1{}^1 \sin\phi). \qquad \ldots\ldots(11)$$

The potentials (9) and (10) will be a solution of the problem if we are able to determine the α's and β's from the boundary conditions. The application of (6) is straightforward. In the case of (7) we use the following procedure:

$$(\epsilon_0 \mathbf{E}_0, \mathbf{n}) = -\epsilon_0 \frac{\partial V_0}{\partial n} = -\epsilon_0 \left(\frac{\partial V_0}{\partial r}\right)_{r=a},$$

$$(\epsilon_1 \mathbf{E}_1, \mathbf{n}) = (\mathbf{n}, \epsilon_1 \mathbf{E}_{1x}) + (\mathbf{n}, \epsilon_1 \mathbf{E}_{1y}) + (\mathbf{n}, \epsilon_1 \mathbf{E}_{1z})$$
$$= (\mathbf{n}, \epsilon_1 \mathbf{i}) E_{1x} + (\mathbf{n}, \epsilon_1 \mathbf{j}) E_{1y} + (\mathbf{n}, \epsilon_1 \mathbf{k}) E_{1z},$$

where **i**, **j**, **k** are unit vectors in the x, y, z directions respectively. These vectors are along the principal axes of ϵ_1 so that

$$(\epsilon_1 \mathbf{E}_1, \mathbf{n}) = (\mathbf{n}, \mathbf{i}) \epsilon_x E_{1x} + (\mathbf{n}, \mathbf{j}) \epsilon_y E_{1y} + (\mathbf{n}, \mathbf{k}) \epsilon_z E_{1z},$$

and since $(\mathbf{i}, \mathbf{n}) = \partial x / \partial r$, etc., we have, using (3) and (8),

$$(\epsilon_1 \mathbf{E}_1, \mathbf{n}) = F(\epsilon_x \beta_1 P_1 + \epsilon_y \beta_2 P_1{}^1 \cos\phi + \epsilon_z \beta_3 P_1{}^1 \sin\phi).$$

With a little algebra, in which we put

$$K_x = \epsilon_x/\epsilon_0, \quad K_y = \epsilon_y/\epsilon_0, \quad K_z = \epsilon_z/\epsilon_0, \qquad \ldots\ldots(12)$$

and make use of the linear independence of the surface harmonics, we obtain finally

$$\beta_1 = \frac{3l}{2+K_x}, \quad \beta_2 = \frac{3m}{2+K_y}, \quad \beta_3 = \frac{3n}{2+K_z}, \qquad \ldots\ldots(13)$$

$$\alpha_1 = -\left(\frac{1-K_x}{2+K_x}\right) a^3 l, \quad \alpha_2 = -\left(\frac{1-K_y}{2+K_y}\right) a^3 m, \quad \alpha_3 = -\left(\frac{1-K_z}{2+K_z}\right) a^3 n, \qquad \ldots\ldots(14)$$

so that the potential problem is solved.

Examination of the potential reveals the following general characteristics of the field. Inside the sphere the field is uniform, as in the analogous isotropic problem, but in this problem the direction of the field is generally different from that of the applied field **F**, the only exceptions being when **F** is directed along a principal axis of ϵ_1. Outside the sphere the field is **F** together with a doublet field. The doublet which, placed at the origin, would produce this part of the field, if the solid sphere were not introduced, is the image doublet of the problem. The components of its moment are

$$\mu_x = -\epsilon_0 F l a^3 \left(\frac{1-K_x}{2+K_x}\right), \quad \mu_y = -\epsilon_0 F m a^3 \left(\frac{1-K_y}{2+K_y}\right), \quad \mu_z = -\epsilon_0 F n a^3 \left(\frac{1-K_z}{2+K_z}\right). \qquad \ldots\ldots(15)$$

We see that the direction of the doublet is generally different from those of both **F** and $\mathbf{E}_1$, exceptions being when **F** is directed along a principal axis of ϵ_1, when all three directions coincide.

We shall now determine the mechanical action on the sphere. The general method is to apply the Maxwell stress tensor at the outer surface, but in this particular problem a more expeditious method is to find the action of the field **F** upon the image doublet, the stresses at radius a with this doublet being the same as those in the actual problem. We thus find that there is no force on the sphere but, after a simple calculation, that there is a couple with components

$$\left.\begin{aligned} G_x &= 3\epsilon_0 F^2 a^3\, mn\left(\frac{1}{2+K_z}-\frac{1}{2+K_y}\right),\\ G_y &= 3\epsilon_0 F^2 a^3\, nl\left(\frac{1}{2+K_x}-\frac{1}{2+K_z}\right),\\ G_z &= 3\epsilon_0 F^2 a^3\, lm\left(\frac{1}{2+K_y}-\frac{1}{2+K_x}\right).\end{aligned}\right\} \qquad \ldots\ldots(16)$$

The existence of a couple is a feature peculiar to the anisotropic problem. There is no mechanical action at all on an isotropic sphere, whereas in the present case the couple is zero only if $\mathbf{F}$ is directed along a principal axis of ϵ_1. The couple is proportional to the volume of the sphere, all other variables having fixed values, and since F enters through the factor F^2, the result is valid for an alternating applied field, provided one takes for F the root mean square value and the frequency is not so high as to necessitate the whole problem being treated electromagnetically.

§ 3. ANISOTROPIC CONDUCTING SPHERE

If the anisotropic sphere and surrounding fluid are uniform linear conductors we have two extra variables to consider, the conductivity σ_0 of the fluid and the symmetric conductivity tensor σ_1 of the solid. The principal axes of σ_1 do not necessarily coincide with those of ϵ_1, and we now choose the coordinate axes to coincide with the principal axes of σ_1 so that this tensor is diagonal:

$$\sigma_1 \equiv \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ & \sigma_{yy} & \sigma_{yz} \\ & & \sigma_{zz} \end{bmatrix} = \begin{bmatrix} \sigma_x & 0 & 0 \\ & \sigma_y & 0 \\ & & \sigma_z \end{bmatrix}, \qquad \ldots\ldots(17)$$

although ϵ_1 generally is not. Apart from our different choice of coordinate axes, the mathematical specification is the same as with the non-conducting case.

The equations which the steady-current potential must satisfy outside and inside the sphere are respectively

$$\nabla^2 V_0 = 0, \qquad \ldots\ldots(18)$$

$$\sigma_x \frac{\partial^2 V_1}{\partial x^2} + \sigma_y \frac{\partial^2 V_1}{\partial y^2} + \sigma_z \frac{\partial^2 V_1}{\partial z^2} = 0 \qquad \ldots\ldots(19)$$

(Smythe 1939), together with the boundary conditions

$$V_0 = V_1 + \lambda \qquad (r=a), \qquad \ldots\ldots(20)$$

$$(\sigma_0 \mathbf{E}_0, \mathbf{n}) = (\sigma_1 \mathbf{E}_1, \mathbf{n}) \qquad (r=a), \qquad \ldots\ldots(21)$$

where λ is a constant.*

We assume solutions similar to (9) and (10), and it will be appreciated at once by comparison of (18)–(21) with (4)–(7) that the analysis is the same as before except that ϵ_0, ϵ_x, ϵ_y, ϵ_z are replaced by σ_0, σ_x, σ_y, σ_z and V_1 is smaller by the amount λ than it would be if we assumed the boundary condition (6). The general characteristics of the field are therefore similar to those in the non-conducting case, although they are determined quantitatively by conductivity rather than dielectric constant, and we note that there exists a distribution of surface charge, this being determined by conductivity and dielectric constant jointly.

---

* In classical steady-current theory the additive constant λ does not appear (e.g. Jeans 1943), but on the basis of modern knowledge inclusion of this constant is necessitated by the existence of a potential ' barrier ' at the surface of a conductor.

The correspondence between σ_0, σ_1 and ϵ_0, ϵ_1 in the two problems fails when we consider the image doublet, for the strength of the doublet must contain a factor to allow for the reduction of electric intensity incurred by the dielectric properties of the fluid. The components of the moment are

$$\mu_x = -\epsilon_0 Fla^3\left(\frac{1-\tau_x}{2+\tau_x}\right),\quad \mu_y = -\epsilon_0 Fma^3\left(\frac{1-\tau_y}{2+\tau_y}\right),\quad \mu_z = -\epsilon_0 Fna^3\left(\frac{1-\tau_z}{2+\tau_z}\right), \qquad \ldots\ldots(22)$$

where we have put

$$\tau_x = \sigma_x/\sigma_0,\quad \tau_y = \sigma_y/\sigma_0,\quad \tau_z = \sigma_z/\sigma_0. \qquad \ldots\ldots(23)$$

Using the image doublet the mechanical action is determined as in the non-conducting case. There is no force on the sphere, but there is a couple with components

$$\left.\begin{aligned} G_x &= 3\epsilon_0 F^2a^3 mn\left(\frac{1}{2+\tau_z}-\frac{1}{2+\tau_y}\right),\\ G_y &= 3\epsilon_0 F^2a^3 nl\left(\frac{1}{2+\tau_x}-\frac{1}{2+\tau_z}\right),\\ G_z &= 3\epsilon_0 F^2a^3 lm\left(\frac{1}{2+\tau_y}-\frac{1}{2+\tau_x}\right).\end{aligned}\right\} \qquad \ldots\ldots(24)$$

This couple shares with (16) certain general properties mentioned, and possesses some in addition. It will be seen that the dielectric constant of the fluid is involved although that of the solid is not.* If $\sigma_0 \to \infty$, then $\mathbf{G} \to 0$. In this limit, however, the problem degenerates into a well-known electrostatic one (Jeans 1943) in which the anisotropy of the solid is unimportant, the only significant feature being that the sphere is a conductor rather than an insulator.

§ 4. ANISOTROPIC NON-CONDUCTING SPHEROID

We now consider the problem of an anisotropic non-conducting spheroid immersed in a non-conducting fluid and influenced by a uniform electric field. In this case it is in general of no advantage to allow the cartesian coordinate axes to coincide with the principal axes of the dielectric tensor as none of these may coincide with the axis of revolution of the spheroid, any simplification gained by a simpler dielectric matrix being offset by a more complex theory of harmonic potentials. We choose as origin the centre of the spheroid and as x-axis the axis of revolution, the y- and z-axes being anywhere in the diametral plane normal to the x-axis. Let a be the semi-axis of revolution and b the other semi-axis. We must distinguish throughout between the cases of a prolate spheroid ($a>b$) and an oblate spheroid ($a<b$). For the former case we define prolate spheroidal coordinates by

$$\left.\begin{aligned} &x = c\xi\eta,\\ &y = c(1-\xi^2)^{1/2}(\eta^2-1)^{1/2}\cos\phi,\\ &z = c(1-\xi^2)^{1/2}(\eta^2-1)^{1/2}\sin\phi,\\ &(-1\leqslant\xi\leqslant+1,\ \eta\geqslant1,\ 0\leqslant\phi<2\pi,\ c=(a^2-b^2)^{1/2}),\end{aligned}\right\} \qquad \ldots\ldots(25)$$

---

* This conclusion might be surprising, for it is intuitively evident that the mechanical action on the surface charge contains ϵ_1, since the surface charge itself does, but an analysis of the total couple in terms of parts with distinct physical origin shows that terms containing ϵ_1 cancel (Cade 1952 b).

and, in the latter case, oblate spheroidal coordinates by

$$\left.\begin{aligned} &x = c\xi\eta\,, \\ &y = c(1-\xi^2)^{1/2}(1+\eta^2)^{1/2}\cos\phi\,, \\ &z = c(1-\xi^2)^{1/2}(1+\eta^2)^{1/2}\sin\phi\,, \\ &(-1 \leqslant \xi \leqslant +1\,,\ \eta \geqslant 0\,,\ 0 \leqslant \phi < 2\pi\,,\ c = (b^2-a^2)^{1/2})\,. \end{aligned}\right\} \qquad \ldots\ldots(26)$$

It will be observed that the meanings of η and c depend upon whether we are treating the prolate or the oblate case, and this is true of other quantities to be introduced later. This fact must be remembered in all subsequent formulae, special care being necessary where one formula applies to both cases. Apart from the symbols just introduced, all those below have the same meaning as in the sphere problems.

As with the non-conducting sphere, the potential problem is to superpose upon the original field that due to the replacement of a fluid spheroid by a solid one with different dielectric properties. Outside the spheroid eqn. (4) must be satisfied and, in accordance with the theory of spheroidal harmonics, we tentatively assume for the respective prolate and oblate cases

$$\begin{aligned} V_o = &-Fc\{lP_1(\xi)P_1(\eta) + m\,P_1^1(\xi)\,P_1^1(\eta)\cos\phi + nP_1^1(\xi)\,P_1^1(\eta)\sin\phi\} \\ &+F\{\alpha_1 P_1(\xi)\,Q_1(\eta) + \alpha_2 P_1^1(\xi)\,Q_1^1(\eta)\cos\phi + \alpha_3 P_1^1(\xi)\,Q_1^1(\eta)\sin\phi\}, \quad \ldots.(27) \end{aligned}$$

$$\begin{aligned} V_o = &+Fci\{lP_1(\xi)\,P_1(i\eta) + mP_1^1(\xi)\,P_1^1(i\eta)\cos\phi + nP_1^1(\xi)\,P_1^1(i\eta)\sin\phi\} \\ &+F\{\alpha_1 P_1(\xi)\,Q_1(i\eta) + \alpha_2 P_1^1(\xi)\,Q_1^1(i\eta)\cos\phi + \alpha_3 P_1^1(\xi)\,Q_1^1(i\eta)\sin\phi\}, \quad \ldots(28) \end{aligned}$$

where the function P_1^1 of imaginary argument or real argument greater than unity is defined by $P_1^1(u) = +(u^2-1)^{1/2}\,d[P_1(u)]/du$ and the α's are constants to be determined. Inside the spheroid (5) must be satisfied, and we make the simple assumption

$$V_i = -\frac{F}{c}(\beta_1 x + \beta_2 y + \beta_3 z)\,, \qquad \ldots\ldots(29)$$

where the β's are constants to be determined. This function is also a solution of Laplace's equation and is expressed in prolate and oblate spheroidal harmonics, respectively, by

$$V_i = -F\{\beta_1 P_1(\xi)\,P_1(\eta) + \beta_2 P_1^1(\xi)\,P_1^1(\eta)\cos\phi + \beta_3 P_1^1(\xi)\,P_1^1(\eta)\sin\phi\}\,, \quad \ldots..(30)$$

$$V_i = +Fi\{\beta_1 P_1(\xi)P_1(i\eta) + \beta_2 P_1^1(\xi)P_1^1(i\eta)\cos\phi + \beta_3 P_1^1(\xi)P_1^1(i\eta)\sin\phi\}. \quad \ldots..(31)$$

The determination of the constants depends upon the dielectric boundary conditions, which are

$$V_o = V_i\,, \qquad (\eta = a/c)\,, \qquad \ldots\ldots(32)$$

$$(\epsilon_o\,\mathbf{E}_o, \mathbf{n}) = (\epsilon_i\,\mathbf{E}_i, \mathbf{n})\,, \quad (\eta = a/c)\,. \qquad \ldots\ldots(33)$$

The application of (32) is straightforward and, using the independence of the Legendre functions, yields one set of three equations for the α's and β's. The application of (33) is more difficult. The right-hand side is dealt with in a similar manner as with the sphere problem, but is more complex on account of the general matrix (1) being used instead of the special form (2). The analysis yields another set of three equations, so that the constants are determinate and the potentials (27)–(31) are an admissible solution of the problem. The values of the constants are

$$\beta_1 = c\frac{\Delta_1}{\Delta}, \qquad \beta_2 = c\frac{\Delta_2}{\Delta}, \qquad \beta_3 = c\frac{\Delta_3}{\Delta}, \qquad \ldots\ldots(34)$$

$$\alpha_1 = \frac{a}{A}\left(l - \frac{\beta_1}{c}\right), \quad \alpha_2 = \frac{b}{B}\left(m - \frac{\beta_2}{c}\right), \quad \alpha_3 = \frac{b}{B}\left(n - \frac{\beta_3}{c}\right), \qquad \ldots\ldots(35)$$

where

$$\left.\begin{aligned}
\Delta_1 = {} & l\left(1 - \frac{aB}{bA}\right)\left[\left(K_{yy} - \frac{bC}{aB}\right)\left(K_{zz} - \frac{bC}{aB}\right) - K_{yz}^2\right] \\
& + \left(1 - \frac{bC}{aB}\right)\left[mK_{xz}K_{yz} + nK_{xy}K_{yz}\right. \\
& \left. - mK_{xy}\left(K_{zz} - \frac{bC}{aB}\right) - nK_{xz}\left(K_{yy} - \frac{bC}{aB}\right)\right], \\
\Delta_2 = {} & \left(1 - \frac{bC}{aB}\right)\left[m\left(K_{zz} - \frac{bC}{aB}\right)\left(K_{xx} - \frac{aB}{bA}\right) + nK_{xy}K_{xz} - mK_{xz}^2\right. \\
& \left. - n\left(K_{xx} - \frac{aB}{bA}\right)K_{yz}\right] - l\left(1 - \frac{aB}{bA}\right)\left[K_{xy}\left(K_{zz} - \frac{bC}{aB}\right)\right. \\
& \left. - K_{xz}K_{yz}\right], \\
\Delta_3 = {} & \left(1 - \frac{bC}{aB}\right)\left[n\left(K_{yy} - \frac{bC}{aB}\right)\left(K_{xx} - \frac{aB}{bA}\right) + mK_{xy}K_{xz} - nK_{xy}^2\right. \\
& \left. - m\left(K_{xx} - \frac{aB}{bA}\right)K_{yz}\right] - l\left(1 - \frac{aB}{bA}\right)\left[K_{xz}\left(K_{yy} - \frac{bC}{aB}\right)\right. \\
& \left. - K_{xy}K_{yz}\right], \\
\Delta = {} & \left(K_{xx} - \frac{aB}{bA}\right)\left[\left(K_{yy} - \frac{bC}{aB}\right)\left(K_{zz} - \frac{bC}{aB}\right) - K_{yz}^2\right] \\
& - K_{xy}\left[K_{xy}\left(K_{zz} - \frac{bC}{aB}\right) - K_{xz}K_{yz}\right] \\
& + K_{xz}\left[K_{xy}K_{yz} - K_{xz}\left(K_{yy} - \frac{bC}{aB}\right)\right],
\end{aligned}\right\} \qquad \ldots\ldots(36)$$

$$K_{st} = \epsilon_{st}/\epsilon_0 \qquad (s, t = x, y, z), \qquad \ldots\ldots(37)$$

and, in the respective prolate and oblate cases,

$$\left.\begin{aligned}
A = \frac{1}{2}\frac{a}{c}\log\frac{a+c}{a-c} - 1, \qquad & B = \frac{1}{2}\frac{b}{c}\log\frac{a+c}{a-c} - \frac{a}{b}, \\
C = \frac{1}{2}\frac{a}{c}\log\frac{a+c}{a-c} & - 2 + \frac{a^2}{b^2},
\end{aligned}\right\} \qquad \ldots\ldots(38)$$

$$\left.\begin{aligned}
A = \frac{a}{c}\cot^{-1}\frac{a}{c} - 1, \qquad & B = \frac{b}{c}\cot^{-1}\frac{a}{c} - \frac{a}{b}, \\
C = \frac{a}{c}\cot^{-1}\frac{a}{c} & - 2 + \frac{a^2}{b^2}.
\end{aligned}\right\} \qquad \ldots\ldots(39)$$

The α's and β's are very complex functions of the matrix elements of ϵ_1 and the constants of the spheroid, but a considerable simplification occurs if one principal axis of ϵ_1 coincides with the axis of revolution, and therefore the x-axis, in which case the y- and z-axes may be chosen to coincide with the other principal axes and ϵ_1 becomes the diagonal matrix (2). By putting

$$\left.\begin{aligned} K_{st} &= 0 \quad (s \neq t), \\ K_{ss} &= K_s \quad (s = x, y, z) \end{aligned}\right\} \qquad \ldots\ldots(40)$$

we obtain from (34)–(36)

$$\beta_1 = lc\left(\frac{1 - aB/bA}{K_x - aB/bA}\right), \qquad \beta_2 = mc\left(\frac{1 - bC/aB}{K_y - bC/aB}\right), \qquad \beta_3 = nc\left(\frac{1 - bC/aB}{K_z - bC/aB}\right), \qquad \ldots\ldots(41)$$

$$\alpha_1 = \frac{a}{A}\, l\left(\frac{K_x - 1}{K_x - aB/bA}\right), \qquad \alpha_2 = \frac{b}{B}\, m\left(\frac{K_y - 1}{K_y - bC/aB}\right), \qquad \alpha_3 = \frac{b}{B}\, n\left(\frac{K_z - 1}{K_z - bC/aB}\right). \qquad \ldots\ldots(42)$$

To determine the mechanical action on the spheroid we begin by considering the components of the Maxwell stress tensor which act upon an element of the outer surface. These are

$$T_{\eta\eta} = \frac{\epsilon_0}{8\pi}(E_\eta^2 - E_\xi^2 - E_\phi^2), \quad T_{\eta\xi} = \frac{\epsilon_0}{4\pi} E_\eta E_\xi, \quad T_{\eta\phi} = \frac{\epsilon_0}{4\pi} E_\eta E_\phi, \qquad \ldots\ldots(43)$$

where E_η, E_ξ, E_ϕ are components of electric intensity at $\eta = a/c$ in the respective directions of η, ξ, ϕ increasing, and are obtained by taking the appropriate gradient of V_o with the aid of the definitions of the spheroidal coordinates, (25) and (26). It is easily found from considerations of spherical symmetry that there is no total force on the spheroid. There is a couple, and this is found by taking the moment of the stress about the centre of the spheroid and integrating over the surface. The calculation is laborious and protracted, although essentially elementary, and the final results for the x- and y-components in the respective prolate and oblate cases are

$$G_x = \tfrac{2}{3}\epsilon_0 F^2 c^2 (m\alpha_3 - n\alpha_2), \qquad \ldots\ldots(44)$$

$$\begin{aligned} G_y = \tfrac{1}{2}\epsilon_0 F^2 \Big\{ \frac{ab^2}{c^2} \Big[\Big(1 - \frac{b^2}{c^2}(A+1)\Big)(la - \alpha_1 A)(na - \alpha_3 C) \\ - A(lb - \alpha_1 B)(nb - \alpha_3 B) \Big] - b(\tfrac{1}{3} - b^2 A/c^2)[(lb - \alpha_1 B)(na - \alpha_3 C) \\ + (la - \alpha_1 A)(nb - \alpha_3 B)] - \tfrac{1}{3} a (lb - \alpha_1 B)(nb - \alpha_3 B) \Big\}, \end{aligned} \qquad \ldots\ldots(45)$$

$$G_x = -\tfrac{2}{3}\epsilon_0 F^2 c^2 (m\alpha_3 - n\alpha_2), \qquad \ldots\ldots(46)$$

$$\begin{aligned} G_y = \tfrac{1}{2}\epsilon_0 F^2 \Big\{ \frac{ab^2}{c^2} \Big[A(nb - \alpha_3 B)(lb - \alpha_1 B) - \Big(1 - \frac{b^2}{a^2}(A+1)\Big)(la - \alpha_1 A) \\ \times (na - \alpha_3 C) \Big] - b(\tfrac{1}{3} + b^2 A/c^2)[(lb - \alpha_1 B)(na - \alpha_3 C) \\ + (la - \alpha_1 A)(nb - \alpha_3 B)] - \tfrac{1}{3} a (lb - \alpha_1 B)(nb - \alpha_3 B) \Big\}. \end{aligned} \qquad \ldots\ldots(47)$$

It is unnecessary to write down the z-components since, by symmetry, these are obtained from the y-components by replacing n, α_3 by m, α_2, respectively, and changing the signs throughout.

It will be appreciated at once that the couples (44)–(47) are extremely complicated functions of the matrix elements of ϵ_1, the direction cosines of $\mathbf{F}$ and the constants of the spheroid. Nevertheless we can extract a few of their general properties from the formulae. In the first place they are homogeneous functions of the third degree in a, b and c, so that for variable a and b, the ratio a/b and all other quantities having fixed values, they are proportional to the volume of the spheroid. Secondly, there is a couple G_x about the axis of revolution, this being a property peculiar to an anisotropic spheroid. It is not surprising that there are non-zero components G_y and G_z as this is so in the isotropic case, but in the isotropic case, for given $\mathbf{F}$, the direction of the couple is determined solely by whether the spheroid is prolate or oblate; in the anisotropic case the direction is specified by the geometry of the spheroid and the tensor ϵ_1 jointly, so that without a knowledge of ϵ_1 one cannot say whether the couple will have the same or the opposite direction as that on a geometrically similar isotropic spheroid. Finally, the applied field enters through the factor F^2, so that the results are valid for an alternating field provided one takes for F the root mean square value and the frequency is not so high as to necessitate the problem being treated electromagnetically *ab initio*.

A great simplification in the formulae occurs in the important case in which the principal axes of ϵ_1 coincide with the cartesian coordinate axes. Using (41) and (42), we find, after some tedious manipulation, that

$$\begin{aligned} G_x = \tfrac{1}{3}\epsilon_0 F^2 ab^2 mn \Bigg[& \left\{ \frac{ab^2}{4c^3}\log\frac{a+c}{a-c} - \frac{a^2}{2c^2} + \frac{1}{1-K_z} \right\}^{-1} \\ & - \left\{ \frac{ab^2}{4c^3}\log\frac{a+c}{a-c} - \frac{a^2}{2c^2} + \frac{1}{1-K_y} \right\}^{-1} \Bigg], \end{aligned} \qquad \ldots\ldots(48)$$

$$\left.\begin{aligned} G_y &= \tfrac{1}{6}\frac{\epsilon_0 F^2 ab^2 nl}{D(K_x, K_z)} \Bigg\{ 2(K_z - K_x) - 3(K_z - 1)(K_x - 1) \\ &\qquad \times \left(\frac{a^2}{c^2} - \frac{2}{3} - \frac{ab^2}{2c^3}\log\frac{a+c}{a-c} \right) \Bigg\}, \\ D(K_x, K_z) &= \left[(K_x - 1)\left(\frac{ab^2}{2c^3}\log\frac{a+c}{a-c} - \frac{b^2}{c^2} \right) + 1 \right] \\ &\qquad \times \left[(K_z - 1)\left(\frac{a^2}{2c^2} - \frac{ab^2}{4c^3}\log\frac{a+c}{a-c} \right) + 1 \right], \end{aligned}\right\} \qquad \ldots\ldots(49)$$

in the prolate case, and

$$\begin{aligned} G_x = \tfrac{1}{3}\epsilon_0 F^2 ab^2 mn \Bigg[& \left\{ \frac{ab^2}{2c^3}\cot^{-1}\frac{a}{c} - \frac{a^2}{2c^2} + \frac{1}{K_y - 1} \right\}^{-1} \\ & - \left\{ \frac{ab^2}{2c^3}\cot^{-1}\frac{a}{c} - \frac{a^2}{2c^2} + \frac{1}{K_z - 1} \right\}^{-1} \Bigg], \end{aligned} \qquad \ldots\ldots(50)$$

$$\left.\begin{aligned} G_y &= \tfrac{1}{6}\frac{\epsilon_0 F^2 ab^2 nl}{D(K_x, K_z)} \Bigg\{ 2(K_z - K_x) + 3(K_z - 1)(K_x - 1) \\ &\qquad \times \left(\frac{b^2}{c^2} - \frac{1}{3} - \frac{ab^2}{c^3}\cot^{-1}\frac{a}{c} \right) \Bigg\}, \\ D(K_x, K_z) &= \left[(K_x - 1)\left(\frac{b^2}{c^2} - \frac{ab^2}{c^3}\cot^{-1}\frac{a}{c} \right) + 1 \right] \\ &\qquad \times \left[(K_z - 1)\left(\frac{ab^2}{2c^3}\cot^{-1}\frac{a}{c} - \frac{a^2}{2c^2} \right) + 1 \right] \end{aligned}\right\} \qquad \ldots\ldots(51)$$

in the oblate case. A noteworthy simplification which has occurred is the way in which the direction cosines of **F** are involved, this being the same as in the case of the sphere (cf. (16)), showing that the couple is now zero if **F** is directed along a principal axis of ϵ_1, which is an axis of the spheroid. From (48)–(51) one may easily obtain simple approximate formulae holding in the extreme cases of a spheroid of large eccentricity ('rod-shaped' in the prolate case and 'disc-shaped' in the oblate case) and one which is nearly spherical. In the latter case one may proceed to the spherical limit when, as is to be expected, formulae (16) reappear.

§ 5. ANISOTROPIC CONDUCTING SPHEROID

In the problem of an anisotropic conducting spheroid in a conducting fluid, and influenced by a uniform electric field, the mathematical specification is the same as in the non-conducting case except that one must include the conductivities of the liquid and solid, the former being a numerical constant σ_0 and the latter the symmetric tensor σ_1.

Remembering the analogy between the problems of the non-conducting sphere and the conducting sphere, it follows, without further comment, that all the formulae given above relating to the non-conducting spheroid hold in the present problem provided that the constants K_{st} are replaced by

$$\tau_{st}=\sigma_{st}/\sigma_0 \qquad (s,\, t=x,\, y,\, z) \qquad \ldots\ldots(52)$$

and the K_s by τ_s, given by (23), and the constant λ is subtracted from the internal potential. It must, of course, be borne in mind that the analogues of (48)–(51) now refer to the principal axes of σ_1, rather than of ϵ_1, coinciding with the coordinate axes.

Since the field is determined by the conductivities, it follows from the joint consideration of conductivity and dielectric constant that the spheroid sustains a charge distribution. There is, in fact, surface charge but no body charge (Cade 1952 a). The main additional points of discussion with regard to the couples are, as with the sphere, that the dielectric constant of the fluid, although not of the solid, is involved, and that the problem has an electrostatic limit in which the only significant property of the spheroid is that it is a conductor rather than an insulator. In this limit G_x is zero, although G_y and G_z are not.

The problem of the couples on an isotropic conducting spheroid occurs as the limiting case of the present one, and has been previously solved by Fürth (1924, 1927). The limit is obtained simply by putting $\tau_x=\tau_y=\tau_z=\tau$. Our result for the prolate case then agrees with Fürth's, and that for the oblate case does when Fürth's is corrected for a small error (Cade 1952 b).

§ 6. POSSIBLE EXPERIMENTAL APPLICATIONS

The main possibility for experimental application of the couples on anisotropic non-conducting spheres and spheroids is with regard to the determination of crystalline dielectric constants. The idea is old and is discussed by Voigt (1910) and Wooster (1949). As mentioned at the beginning, however, the theoretical background to the method, as given by these authors, is unsatisfactory. We consider briefly the sources of error.

In the first place Voigt and Wooster treat the analogous paramagnetic problem, in which the internal field is assumed to differ negligibly from the external field, and then carry the results over to the dielectric case. Since, however, electric susceptibilities are usually of a higher order of magnitude than paramagnetic susceptibilities, the above-mentioned assumption is generally invalid for a dielectric; instead, the exact determination of the field is necessary. Secondly, it is assumed that if the internal field is uniform, the only mechanical action is a couple density of the form

$$\mathbf{g}=[\mathbf{P},\ \mathbf{E}], \qquad \ldots\ldots(53)$$

where **P** is the dielectric polarization. This assumption is wrong since it overlooks the non-uniformity of the field at the dielectric boundary and since Brown's hydrostatic stresses are omitted. The couple (53) is, in fact, just one of a set of terms which lead ultimately to the Maxwell stresses, which we have used. Finally, the effect of a fluid dielectric environment is wrongly treated, it being assumed that the introduction of the dielectric necessitates the transformation

$$\epsilon_s-1\rightarrow\epsilon_s-\epsilon_0 \qquad (s=x,\,y,\,z). \qquad \ldots\ldots(54)$$

It is shown elsewhere (Cade 1952 b) that, with the cancellation of this erroneous transformation, the last two errors amount together to the omission of a normal surface traction on the immersed body.

We conclude that the significant contributions of the present work on non-conducting media to crystal physics are the exact determination of the electrostatic field in special cases and the rigorous and critical application of mechanical action theory.

The formulae assumed by Voigt and Wooster for the couple on a crystal of any shape, in empty space and influenced by a uniform electric field, are

$$\left.\begin{aligned} G_x&=\frac{mn}{4\pi}vF^2(\epsilon_y-\epsilon_z),\\ G_y&=\frac{nl}{4\pi}vF^2(\epsilon_z-\epsilon_x),\\ G_z&=\frac{lm}{4\pi}vF^2(\epsilon_x-\epsilon_y),\end{aligned}\right\} \qquad \ldots\ldots(55)$$

where v is the volume of the crystal. Although this result is in general wrong, our general analysis of the fundamental errors shows that these are ineffectual as regards components of **G** about axes of revolution, if such exist, if $K_x\simeq K_y\simeq K_z\simeq 1$. With the last-mentioned restriction we would therefore expect out formulae (16), (48) and (51) to approximate to (55), and this is found to be the case.

Voigt and Wooster found the mechanical action method to be an unsatisfactory one for the determination of dielectric tensors, and it now appears that this conclusion might be the result of the errors of theory which led to the formulae (55). For example, it was found that the couple was sometimes in the opposite direction from what was expected. This was attributed to 'charge leakage', but upon examination of the couple (49) for the prolate spheroid, for example, we easily find that the second term in the curved bracket could more than outweigh the first, and a term of the form of the first is all that appears on Voigt's and Wooster's ideas of the mechanical action. We therefore consider that the method might be revived on the basis of the present theory. It has an advantage over the standard method, which is to use a crystal plate as dielectric in a condenser. According to Wooster (1949), one needs a plate of linear dimensions "a few millimetres". Much smaller crystals could be used with the mechanical action method; the only limitation is the fineness of suspension obtainable.

We suggest that one should use either a spherical crystal or a spheroidal one suspended with its axis of revolution vertical. A twofold advantage is derived by this choice. In the first place one avoids the mechanical action controversy; it has been shown elsewhere (Cade 1952 b) that all theories of mechanical action lead to formulae (16), (44), (46), (48) and (50), whereas this is not the case with (45), (47), (49) and (51). Secondly, the couple is due solely to the anisotropy. In the case of (45), (47), (49) and (51) part of the couple arises from the geometry of the spheroid (there is a couple on a geometrically similar isotropic spheroid), so that there would be a greater experimental error in the determination of constants characterizing the anisotropy.

In the case of couples on conducting crystals, we suggest that the results be used for the determination of crystalline conductivities. Such a method would be completely new, since, as previously mentioned, the steady-current problems have not been solved before. According to Wooster (1949), the only method hitherto used has been the simple determination of the resistance of single crystal wires. This method only works with metals, where single crystal wires can be made. On the other hand, our method would be useless for metals since, because metallic conductivities are invariably large, it is found from consideration of the relevant formulae that it is experimentally impossible to obtain couples of appreciable magnitude. Our method would be of value in the case of crystals of salts, for example, where single crystal wires are unobtainable and where, indeed, no method appears previously to have been used. We conclude that the two methods are complementary, and that the one we propose would add a large new domain to the technology of crystalline conductivity determinations.

ACKNOWLEDGMENT

The author is indebted to Dr. R. Fürth of Birkbeck College, University of London, for suggesting the problems and for valuable advice given at various stages of the work.

REFERENCES

ABRAHAM, M., and BECKER, R., 1932, *Electricity and Magnetism* (London : Blackie).
BROWN, W. F., Jr., 1951, *Amer. J. Phys.*, **19**, 290.
CADE, R., 1951, *Proc. Phys. Soc.* A, **64**, 665; 1952 a, *Ibid.*, **65**, 287; 1952 b, *Ph.D. Thesis*, University of London.
FÜRTH, R., 1924, *Z. Phys.*, **22**, 98; 1927, *Ibid.*, **44**, 256.
JEANS, J., 1943, *Electricity and Magnetism* (Cambridge : University Press).
LARMOR, J., 1897, *Phil. Trans. Roy. Soc.* A, **190**, 280.
LIVENS, G. H., 1926, *The Theory of Electricity* (Cambridge: University Press).
SMITH-WHITE, W. B., 1949, *Phil. Mag.*, **40**, 466.
SMYTHE, W. R., 1939, *Static and Dynamic Electricity* (New York : McGraw-Hill).
VOIGT, W., 1910, *Lehrbuch der Kristallphysik* (Leipzig : Teubner).
WOOSTER, W. A., 1949, *Crystal Physics* (Cambridge : University Press).

Electro-Thermal Behaviour of Point Contacts to Semiconductors

By A. D. STUCKES

Research Department, Metropolitan-Vickers Electrical Co. Ltd., Manchester

Communicated by B. G. Churcher; *MS. received* 9th *January* 1953, *and in final form* 9*th April* 1953

Abstract. The heating effects that occur when a current is passed through a point contact to a semiconductor which has no barrier layer, have been investigated. The d.c. volt–current curve has been calculated for three different activation energies, 0·60, 0·48 and 0·34 ev of the semiconductor with a tungsten contact. The change with frequency of the impedance of the contact to a small sinusoidal ripple, superimposed upon a steady current, has been calculated for an activation energy 0·60 ev. Both these calculations provide a means of estimating the radius of the contact. Experimental d.c. and a.c. characteristics have been obtained for tungsten in contact with discs of sintered magnesium titanate (Mg_2TiO_4) of appropriate activation energies. Fairly good agreement is obtained between the values of the contact radius calculated from the d.c. and a.c. experiments, and between the calculated and measured variation with frequency of the impedance to the sinusoidal ripple. To explain minor anomalies in the d.c. characteristic it is assumed that field emission occurs around the contact.

§ 1. INTRODUCTION

THE volt–current characteristic of a point contact to a semiconductor is, in general, non-linear and may or may not be symmetrical about zero. The three main effects involved in point contact problems are Joule heating due to the current flowing, barrier layer effects and field emission around the contact area.

Heating effects can cause non-linearity, for as the semiconductor has a negative temperature coefficient of resistance, the current will increase more rapidly than the voltage as soon as it is great enough to cause appreciable heating. If there is a sufficiently high rate of change of resistance with temperature and sufficiently small heat loss the voltage will decrease as the current increases and the steady state characteristic will display a negative slope resistance (fig. 1). These effects are essentially time dependent, but occur rapidly enough to be important even when high frequencies and short time impulses are employed. The heating will also give rise to thermoelectric potentials, but these are relatively small and have been neglected in this paper.

Barrier layer effects include thermionic emission of electrons over a potential barrier at the surface of the semiconductor surface and transmission through it. They produce asymmetrical volt–current curves at a metal semiconductor contact when the applied potential exceeds the order of $\frac{1}{10}$ volt. They will always cease to be important when the total voltage is so high that a significant contribution from the barrier layer would imply a field strength much above 10^7 v cm^{-1} within it. (In most cases this will be an overestimate of the limit.)

Field emission does not occur identically from different materials, and therefore gives rise to rectification at metal semiconductor contacts. It requires a field strength at the emitting surface of about 10^7 v cm^{-1} and its occurrence around a contact will only be noticeable if it takes place over an area comparable with that of the contact itself. With any probable shape of metal point this is unlikely to happen when the total potential is much below the order of 100 volts.

It therefore seems reasonable to assume that if the volt–current curve is symmetrical about zero at low voltages then barrier layer effects are negligible. This is the case with many oxide semiconductors, presumably because the density of impurity centres is high, and the space-charge barrier is therefore too thin to produce a potential drop comparable with that caused by the spreading resistance within the body of the material. Examples of this are TiO_2, CuO, Fe_3O_4, Mg_2TiO_4 and NiO. From measurements of the loss of weight during reduction of Mg_2TiO_4 (Jones 1950, unpublished) and from the analysis of substituted lithium in NiO (Mitchell 1952) the density of impurity centre appears to be about 10^{21}, when the conductivity is about 10^{-1} or 10^{-2} $(\Omega\mathrm{cm})^{-1}$ the space-charge barrier layer thickness would be of the order of 10^{-7} cm; this is equivalent only to a single layer of impurities, so the barrier will be both very thin and very patchy.

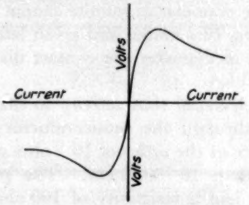


Fig. 1. Steady state voltage–current curve for metal–semiconductor contact with negligible space-charge barrier.

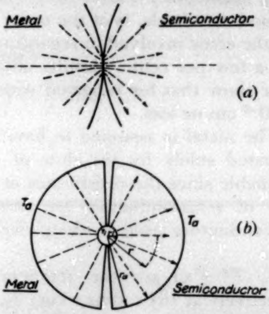


Fig. 2. Model of metal–semiconductor contact.

Heating effects will always occur and must therefore modify the characteristic where a barrier layer (or field emission) is present. This can be seen in the volt–current curves at a point contact to germanium, silicon or lead sulphide; in each case the back characteristic has a marked region of negative slope, which appears from impulse measurements (Benzer 1949, Bennett and Hunter 1950) and from high frequency observations of the author to be predominantly caused by heating. Where the barrier is not very thick its reverse resistance will become comparable with the spreading resistance of the material behind it as soon as it gets hot, and conditions are then similar to those where no barrier exists.

The calculations for a contact with no barrier layer were carried out to see whether the observed behaviour of contacts believed to be of this type could be accounted for and to see whether the observations would give reliable information about the size of the contact and the properties of the semiconductor. It was also considered that this would be a useful step towards understanding the behaviour of some materials (such as silicon carbide) where barrier layer effects are important but do not explain a large part of the observed behaviour. Various phenomena associated with small area contacts to semiconductors may be attributed to self-heating of the contact point, for example the restriking of a welding arc through a layer of semiconducting slag, and the theory provides various tests of the plausibility of such an explanation.

§2. THEORY

The simplest configuration of a small area contact between a metal and a semiconductor would be that shown in fig. 2 (*a*). This has been idealized to the model shown in fig. 2 (*b*) which was used by Amrein (1942) in an earlier treatment of the same problem. It consists of a hemisphere of metal connected to a hemisphere of semiconductor by a small sphere, which is supposed to have infinite electrical and thermal conductivity and zero heat capacity, and of diameter equal to the actual contact diameter. This is a somewhat artificial model but introduces spherical symmetry into the problem thereby simplifying the equations.

In applying the results from this model the hemispheres become semi-infinite blocks; although this may at first sight seem unreasonable for the metal point, elementary mechanical considerations show that the contact diameter, for moderate loads, is of the order of 10^{-4} cm for a wire 0·01 cm in diameter, so that the error involved in regarding the wire diameter as infinite cannot be more than a few per cent. Results obtained using this model and given later in the paper show that for tungsten wire 0·01 cm in diameter the contact diameter is 2×10^{-4} cm or less.

The metal is assumed to have infinite electrical conductivity so that heat is generated solely by the flow of current through the semiconductor; this is reasonable since the resistivities of metals are of the order of 10^{-5} ohm cm while those of semiconductors are rarely less than 10^{-2} ohm cm. The particular semiconductors used in these experiments had a resistivity of 100 ohm cm or more.

T, T^* (°K) are the temperatures within the semiconductor and metal respectively at any radius r cm; T_i, T_o (°K) are the temperatures at the inner and outer boundaries respectively (the same for both semiconductor and metal); r_i, r_o are the radii of the inner and outer boundaries; λ, λ^* (w cm^{-1} °C) are the thermal conductivities of the semiconductor and metal respectively; C, C^* (joules cm^{-3}) are the thermal capacities per unit volume for the semiconductor and metal respectively. The resistivity of the semiconductor is taken as $\rho=\rho_\infty\,\mathrm{e}^{E/2\boldsymbol{k}T}=\rho_\infty\mathrm{e}^{B/T}$ where $B=E/2\boldsymbol{k}$ and ρ_∞ is a constant. E is the activation energy of the semiconductor, and $\boldsymbol{k}$ is Boltzmann's constant.

The voltage V across the semiconductor for a current density j is

$$V=-\int_{r_1}^{r_0} dV=\int_{r_1}^{r_0} j\rho\,dr=\int_{r_1}^{r_0}\frac{J}{4\pi r^2}\rho_\infty\,\mathrm{e}^{B/T}\,dr \qquad \ldots\ldots(1)$$

where $\frac{1}{2}J$ is the total current flowing through the contact.

2.1. *Steady State Condition*

The steady state equation for the temperature within the semiconductor or the metal is

$$\operatorname{div}(\lambda\operatorname{grad}T)+j(-\operatorname{grad}V)=0 \qquad \ldots\ldots(2)$$

giving

$$\lambda\left\{\frac{d^2T}{dr^2}+\frac{2}{r}\frac{dT}{dr}\right\}+\rho\left(\frac{J_s}{4\pi r^2}\right)^2=0 \qquad \ldots\ldots(3)$$

in the semiconductor, and

$$\lambda^*\left\{\frac{d^2T^*}{dr^2}+\frac{2}{r}\frac{dT^*}{dr}\right\}=0 \qquad \ldots\ldots(4)$$

in the metal where $\frac{1}{2}J_s$ is the steady current flowing.

These have to be solved subject to the boundary conditions

$$T=T^*=T_0 \text{ at } r=r_0=\infty \text{ (outer boundary)} \qquad \ldots\ldots(5)$$

$$\lambda\left(\frac{dT}{dr}\right)_{r=r_1}=-\lambda^*\left(\frac{dT^*}{dr}\right)_{r=r_1} \text{ (inner boundary).} \qquad \ldots\ldots(6)$$

Substituting $\rho=\rho_\infty\,e^{B/T}$ and $r=(\rho_\infty/\lambda)^{1/2}J_s/4\pi x$ eqn. (3) for the semiconductor reduces to

$$\frac{d^2T}{dx^2}+e^{B/T}=0 \qquad \ldots\ldots(7)$$

which has the solution

$$x=\int_{T_0}^{T}\left\{\left(\frac{dT}{dx}\right)^2_{x=0}-2\int_{T_0}^{T}e^{B/T}\,dT\right\}^{-1/2}dT \qquad \ldots\ldots(8)$$

fulfilling the boundary condition (5). The sign of the root is to be taken as the same as the sign of the expression inside it.

The solution of eqn. (4) for the metal is simply

$$\frac{T^*-T_0}{T_1-T_0}=\left(\frac{1}{r}-\frac{1}{r_0}\right)\Big/\left(\frac{1}{r_1}-\frac{1}{r_0}\right)=\frac{r_1}{r} \text{ when } r_0=\infty.$$

This gives

$$\left(\frac{dT^*}{dr}\right)_{r=r_1}=-\frac{T_1-T_0}{r_1}$$

so that the boundary condition (6) becomes

$$\left(\frac{dT}{dr}\right)_{r=r_1}=\frac{\lambda^*}{\lambda}\frac{T_1-T_0}{r_1}$$

or substituting as before for r

$$\left(\frac{dT}{dx}\right)_{x=x_1}=-\frac{\lambda^*}{\lambda}\frac{T_1-T_0}{x_1}. \qquad \ldots\ldots(9)$$

Returning to eqn. (1) for the potential between the inner and outer boundaries of the semiconductor, substituting x for r gives

$$V=(\rho_\infty\lambda)^{1/2}\int_{x_0}^{x_1}e^{B/T}\,dx=-(\rho_\infty\lambda)^{1/2}\int_{x_0}^{x_1}\frac{d^2T}{dx^2}\,dx \qquad \ldots\ldots(10)$$

or

$$\frac{V}{(\rho_\infty\lambda)^{1/2}}=\left(\frac{dT}{dx}\right)_{x_0}-\left(\frac{dT}{dx}\right)_{x_1}. \qquad \ldots\ldots(11)$$

Equation (8) has been integrated numerically with $T_0=293^\circ\text{K}$ and with three values of B, namely 3500, 2800 and 2000 (corresponding to activation energies 0·60, 0·48 and 0·34 ev), starting in each case with various values of $(dT/dx)_{x_0}$ covering a range sufficient for experimental conditions. The resulting curves of T and dT/dx plotted against x for $B=3500$ are shown in figs. 3 and 4. One point satisfying condition (9) occurs on each of the (T, x) curves where the right-hand arm is falling towards T_0. The complete (T, x) curve is symmetrical however, so that it is unnecessary to plot the whole of it once x_m the point of maximum T is found. These curves can be used to find the steady state behaviour of a small contact at any metal to a semiconductor having an activation energy in the specified range (or of a contact between two semiconductors) at an ambient temperature of 20°c.†

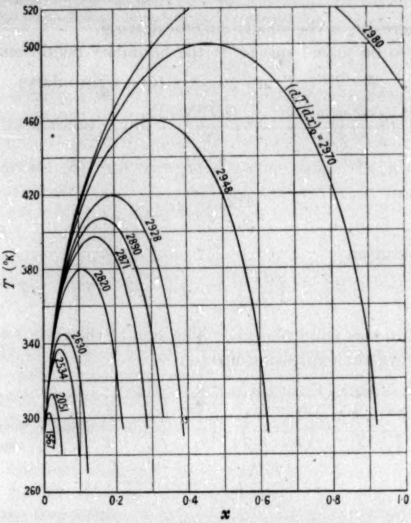


Fig. 3. Graph of T against x for successive values of $(dT/dx)_0$ for $B=3500$.

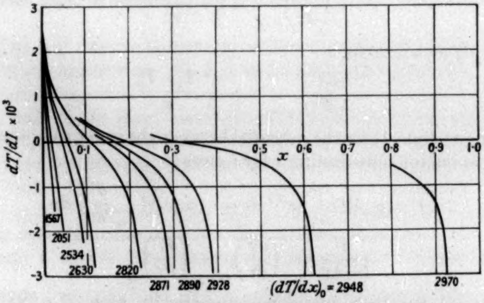


Fig. 4. Graph of dT/dx against x for successive values of $(dT/dx)_0$ for $B=3500$.

A series of values of $(dT/dx)_{x_1}$, T_1 and x_1 for the corresponding values of $(dT/dx)_{x_0}$ satisfying eqn. (9) with $\lambda^*/\lambda=40$ have been obtained in this way;

---

† It can be seen, by putting eqns. (8)–(11) in a slightly more general form using T/B and $x/B^{1/2}$ as variables instead of T and x, that the results of these computations can be applied to all values of T_0 and B such that T_0/B is unaltered. As the present work is only concerned with one ambient temperature, however, the results have been expressed in terms of T and x as being easier to appreciate physically. They apply to any other ambient temperature T_0' with all values of T altered in the ratio $T'=TT_0'/293$, and with $B'=BT_0'/293$, $x'=x(T_0'/293)^{1/2}$ and $V'=V(T_0'/293)^{1/2}$.

hence $V(\rho_\infty\lambda)^{-1/2}$ has been evaluated as a function of x, i.e. as a function of $(\rho_\infty/\lambda)^{1/2}I_s/2\pi r_1$ where I_s is the steady current flowing. As the constant ρ_∞ has little practical use, these results have been expressed in terms of $V(\rho_0\lambda)^{-1/2}$ and $(\rho_0/\lambda)^{1/2}I_s/2\pi r_1 = x_1'$, where ρ_0 is the resistivity at temperature T_0 (293°K), and are plotted in fig. 5 for the three values of B; these are, in effect, voltage–current curves.

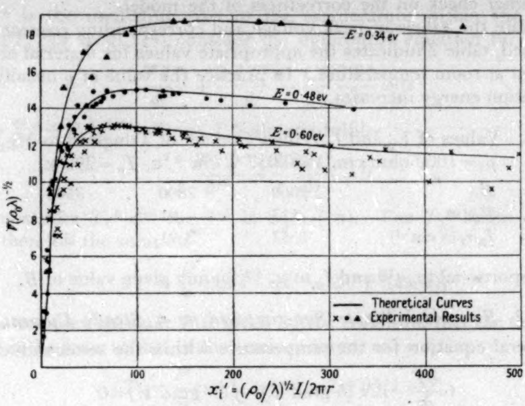


Fig. 5. Voltage–current curves for a tungsten contact to semiconductors of 0·60, 0·48, and 0·34 ev activation energy.

The most easily identifiable points on these curves are the peaks, that is the voltage maxima. Values of the parameters in fig. 5 at the respective peaks (indicated by the suffix p) are set out in the third and fourth rows of table 1.

Table 1

E (ev)	0·34	0·48	0·60
B	2000	2800	3500
$[V(\rho_0\lambda)^{-1/2}]_p$	18·7	15	13
$(x_1')_p$	170	100	78·5
$K = r_1/V_pI_p(\times 10^{-3})$	1·36	2·88	4·20
R_0I_p/V_p	6·9	5·2	3·2

The values of $[V(\rho_0\lambda)^{-1/2}]_p$ can be used to find ρ_0 from the peak voltage if an approximate value of B is known. By multiplying the two parameters together however, and remembering $r_1 = (\rho_0/\lambda)^{1/2}I_p/2\pi x_1$ we get

$$r_1 = \frac{V_pI_p}{2\pi\lambda(x_1')_p V(\rho_0\lambda)^{-1/2}} = KV_pI_p$$

where K depends only on λ, T_0 and B and has the values shown in table 1. The factor K varies linearly with B and as the variation is not rapid, r_1 can be found with fair approximation even if B is not known exactly.

The resistance of the contact for currents not large enough to cause appreciable heating is, on the model of fig. 2,

$$R_0 = \frac{\rho_0}{2\pi r_1} = \rho_\infty \frac{\exp B/T_0}{2\pi r_1}$$

and by combining this with the previous results one finds that the ratio of this initial resistance to the resistance at the peak voltage is given by

$$R_o I_p / V_p = (x_1')_p / [V(\rho_o \lambda)^{-1/2}]_p$$

which depends only on T_o, B and λ, and is shown in table 1. This relation should provide another check on the correctness of the model.

To indicate the values of peak voltage and corresponding current which are to be expected, table 2 indicates the appropriate values for material of resistivity 1000 ohm cm at room temperature. In practice the value of ρ usually increases as the activation energy increases.

Table 2. Values of V_p and I_p/r_1 for a Contact of Tungsten to Mg_2TiO_4
$\rho_o = 1000$ ohm cm, $\lambda = 0{\cdot}037$ w cm^{-1} °c, $T_o = 293$°K

B	2000	2800	3500
V_p (v)	113	91	79
I_p/r_1 (A cm^{-1})	6·53	3·91	3·0

V_p is proportional to $\rho_o^{1/2}$ and I_p to $\rho_o^{-1/2}$ for any given value of B.

2.2. *Sinusoidal Ripple Superimposed on a Steady Current*

The general equation for the temperature within the semiconductor is

$$C\frac{\partial T}{\partial t} - \operatorname{div}(\lambda \operatorname{grad} T) - j(-\operatorname{grad} V) = 0$$

and in spherical coordinates is

$$\frac{\partial^2 T}{\partial r^2} + \frac{2}{r}\frac{\partial T}{\partial r} - \frac{C}{\lambda}\frac{\partial T}{\partial t} + \frac{\rho}{\lambda}\left(\frac{J}{4\pi r^2}\right)^2 = 0. \qquad \ldots\ldots(12)$$

This equation is extremely difficult to solve where J is a sinusoidal function. It can be solved for a steady current J_s flowing through the contact with a superimposed small sinusoidal ripple $J_1 \mathrm{e}^{i\omega t}$:

$$J = J_s + J_1 \mathrm{e}^{i\omega t} \quad \text{where} \quad J_1 \ll J_s$$

so that we can assume with sufficient accuracy

$$T = T_s + T_1 \mathrm{e}^{i\omega t}, \qquad J^2 = J_s^2 + 2J_1 J_s \mathrm{e}^{i\omega t}$$

and

$$\exp\left(\frac{B}{T_s + T_1 \mathrm{e}^{i\omega t}}\right) = \left\{1 - \frac{BT_1}{T_s^2}\mathrm{e}^{i\omega t}\right\} \exp(B/T_s). \qquad \ldots\ldots(13)$$

Substituting these approximations in eqn. (12) and also writing as before $r = (\rho_\infty/\lambda)^{1/2} J_s / 4\pi x$,

$$\frac{d^2 T_s}{dx^2} + \frac{d^2 T_1}{dx^2}\mathrm{e}^{i\omega t} - \frac{Ci\omega\rho_\infty J_s^2}{x^4 \lambda^2 (4\pi)^2} T_1 \mathrm{e}^{i\omega} + \left[1 - \frac{BT_1}{T_s}\mathrm{e}^{i\omega t} + \frac{2J_1}{J_s}\mathrm{e}^{i\omega t}\right] \exp(B/T_s) = 0$$

but the steady state condition $T_1 = 0$, $J_1 = 0$ gives $d^2 T_s/dx^2 + \exp(B/T_s) = 0$ hence

$$\frac{d^2 T_1}{dx^2} - \left\{\frac{iA}{x^4} + \frac{B}{T_s^2}\exp(B/T_s)\right\} T_1 + \frac{2J_1}{J_s}\exp(B/T_s) = 0 \qquad \ldots\ldots(14)$$

where $A = C\omega\rho_\infty J_s^2 / \lambda^2 (4\pi)^2$.

The boundary conditions are the same as those of eqns. (5) and (6) and since the steady part of the temperature already obeys these, we are left with $T_1=0$ at $r=\infty$ or $x=0$, and

$$\lambda\left(\frac{\partial T_1}{\partial r}\right)_{r=r_1} = -\lambda^*\left(\frac{\partial T_1{}^*}{\partial r}\right)_{r=r_1} \qquad \ldots\ldots(15)$$

where $T_1{}^*$ is the oscillating part of the temperature within the metal. To eliminate $\partial T_1{}^*/\partial r$ from (15) consider the differential equation for heat flow in the metal:

$$\frac{\partial^2 T^*}{\partial r^2} + \frac{2}{r}\frac{\partial T^*}{\partial r} - \frac{C^*}{\lambda^*}\frac{\partial T^*}{\partial t} = 0 \qquad \ldots\ldots(16)$$

with $T^* = T_s{}^* + T_1{}^*\mathrm{e}^{i\omega t}$. Since $T_s{}^*$ already fulfils

$$\frac{\partial^2 T_s{}^*}{\partial r^2} + \frac{2}{r}\frac{\partial T_s{}^*}{\partial r} = 0$$

we can substitute $T_1{}^*\,\mathrm{e}^{i\omega t}$ for T^* in eqn. (16). For $T_1{}^*$ finite at $r=\infty$ this equation then has the solution

$$T_1{}^* = \frac{P}{r}\exp\{-r(C^*\omega/2\lambda^*)^{1/2}(1+i)\}$$

where the complex constant P is given by the condition that the temperature of the metal is the same as that of the semiconductor at $r=r_1$, i.e.

$$T_{1i} = \frac{P}{r_1}\exp\{-r_1(C^*\omega/2\lambda^*)^{1/2}(1+i)\}.$$

This leads to

$$\left(\frac{\partial T_1{}^*}{\partial r}\right)_r = -T_{1i}\left\{\frac{1}{r_1} + \left(\frac{C^*\omega}{2\lambda^*}\right)^{1/2}(1+i)\right\}$$

and eqn. (15) becomes

$$\left(\frac{\partial T_1}{\partial r}\right)_{r=r_1} = \frac{\lambda^*}{\lambda}\frac{T_{1i}}{r_1}\left\{1 + r_1\left(\frac{C^*\omega}{2\lambda^*}\right)^{1/2}(1+i)\right\}$$

or

$$\left(\frac{\partial T_1}{\partial x}\right)_{x=x_1} = -\frac{\lambda^*}{\lambda}\frac{T_{1i}}{x_1}\left\{1 + \frac{1+i}{x_i}\left(\tfrac{1}{2}A\frac{C^*\lambda}{C\lambda^*}\right)^{1/2}\right\} \qquad \ldots\ldots(15\,a)$$

where A is the same quantity as in eqn. (14). So far as concerns the experiments in this paper, the second term in the bracket in eqn. (15 *a*), after substituting the appropriate values for x_i, C, C^*, λ and λ^* reduces to $0{\cdot}15\,A^{1/2}(1+i)$ and was therefore neglected for A not greater than 10^{-2}.

The boundary conditions then become $T_1=0$ at $r=\infty$

$$\left(\frac{\partial T_1}{\partial x}\right)_{x=x_1} = -\frac{\lambda^*}{\lambda}\frac{T_{1i}}{x}. \qquad \ldots\ldots(15\,b)$$

Equation (14) has been solved with these boundary conditions using relaxation methods. The steady state solution chosen was that corresponding to $x_i=0{\cdot}378(x_i'=148)$ in the case where $B=3500$, $T_0=293$, $\lambda^*/\lambda=40$, that is the case to which fig. 3 and the lower curve of fig. 5 refer.

The total voltage between the inner and outer boundaries of the semiconductor is still given by eqn. (11), which on splitting J and T into stationary and varying components gives

$$V = \int_{r_1}^{r_0}\frac{J_s + J_1\,\mathrm{e}^{i\omega t}}{4\pi r^2}\,\rho_\infty\exp\left(\frac{B}{T_s + T_1\mathrm{e}^{i\omega t}}\right)dt.$$

Using eqn. (13) and neglecting the second order terms in T_1J_1 this becomes

$$V=\int_{r_1}^{r_0}\frac{\rho_\infty \exp(B/T_s).\{J_s+(J_1-J_sBT_1/T_s^2)\,\mathrm{e}^{i\omega t}\}}{4\pi r^2}\,dr=V_s+V_1\,\mathrm{e}^{i\omega t}$$

where V_s is the steady component of voltage corresponding to J_s and calculated in the previous section:

$$V_1\frac{J_s}{J_1}=V_s-2(\rho_\infty\lambda)^{1/2}\int_{x_0}^{x_1}\frac{BT_1'}{T_s^2}\exp(B/T_s)\,dx \qquad \ldots\ldots(17)$$

and

$$T_1'=T_1J_s/2J_1.$$

The real and imaginary parts of T_1' have been plotted in fig. 6; the real and imaginary parts of $V_1J_s/J_1(\rho_\infty\lambda)^{1/2}$ have been evaluated from these curves.

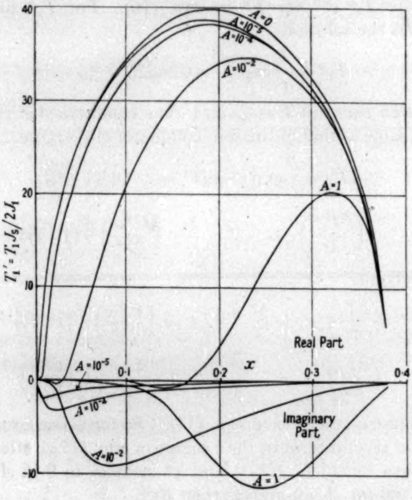


Fig. 6. Oscillatory component of temperature plotted against x.

If we write I_s for the steady current through the contact and Z for the impedance presented to the ripple, $I_s=\frac{1}{2}J_s$, $Z=2V_1/J_1$ so that

$$V_1J_s/J_1=I_sZ=I_s(R_c+iX_c).$$

The quantities $I_sR_c(\rho_0\lambda)^{-1/2}$ and $I_sX_c(\rho_0\lambda)^{-1/2}$ are plotted in fig. 7 against log A', where $A'=C\omega\rho_0I_s^2/(2\pi\lambda)^2$ and is directly proportional to the frequency. The real part of the impedance, the resistive component R_c, is first negative, then zero, then positive as the frequency increases. The inductive component X_c rises from zero through a maximum and then decreases with increasing frequency.

The critical frequency at which the resistive component is zero is easily found experimentally and this provides a second means of calculating the radius of the contact. For, from fig. 7 this frequency is such as to give $A' = 4{\cdot}1$, or since

$$A' = \frac{C\rho_0 I_s^2 \omega}{(2\pi\lambda)^2} = \frac{C\omega r_1^2 (x_1')^2}{\lambda} \quad \text{and} \quad x_1' = 148, \qquad \ldots\ldots(18)$$

hence

$$r_1 = \left(\frac{0{\cdot}3778 \times 10^{-7}}{f_{\text{crit}}}\right)^{1/2} \qquad \ldots\ldots(19)$$

with $\omega = 2\pi f$, $C = 3{\cdot}0$ joules cm^{-3} °c (0·72 cal cm^{-1} sec^{-1} °c) and $\lambda = 0{\cdot}037$ w cm^{-1} °c.

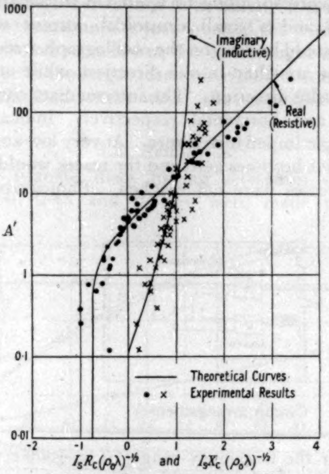


Fig. 7. Impedance of contact to a small sinusoidal ripple against frequency.

§ 3. EXPERIMENT

3.1. *Procedure*

A tungsten contact made from 0·01 cm diameter wire was etched in fused sodium nitrite and then held with firm spring pressure against the surface of a sintered disc of magnesium titanate Mg_2TiO_4. A large area contact was made by brass sprayed on to the other side of the disc. Bulk measurements of resistivity and temperature coefficient of resistance had previously been made on the discs, which were chosen to have B approximately equal to 3500, 2800 or 2000. The thermal conductivity of the magnesium titanate is approximately 0·037 w cm^{-1} °c and that of tungsten 1·47 w cm^{-1} °c and the thermal capacity per unit volume of the titanate and tungsten respectively is 3·0 joules cm^{-3} °c and 2·7 joules cm^{-3} °c.

The circuit shown in fig. 8 was used to measure the d.c. voltage–current curves and to delineate the ripple voltage–current relation on the cathode-ray oscillograph. The 20-megohm resistance in series with the power pack was necessary to limit the current and enable the negative resistance part of the characteristic to be followed. The 2000 ohm resistance in the vertical amplifier lead was necessary to prevent oscillation when the point contact to the semiconductor was in a state of negative slope resistance. The amplifiers were those of a Dumont oscillograph type 208 and were found to introduce negligible phase shift from 0 to 3000 c/s. Owing to the high impedance of the point contact, measurements could not be relied upon, in any case, at higher frequencies.

Measurements of direct currents and voltages across the contact in the steady state were made on discs covering the range of values of B from 2000 to 3500. In the case of discs with $B=3500$, a current equal to 1·9 times the steady current at the peak voltage (corresponding to $x_1=0{\cdot}378$ in the theory) was then passed through the contact and a small sinusoidal current superimposed upon it. In general an ellipse could be seen on the oscillograph screen. At low frequencies the axes of the ellipse are tilted in one direction, while at high frequencies they are tilted in the opposite direction. The intermediate condition, where the axes are nearly vertical and horizontal respectively, indicates that the resistive component of the ripple impedance is zero. At very low and very high frequencies the reactive component becomes zero and the traces would become tilted straight lines, but these extremes were not reached. Photographs of the oscillograph trace were taken over the frequency range 10 to 3000 c/s and a typical set are shown in fig. 10 (Plate). The point contact was then replaced by a standard carbon resistance and another set of photographs taken in order to calibrate the amplifiers.

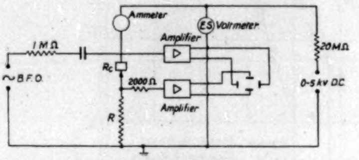


Fig. 8. Circuit arrangement.

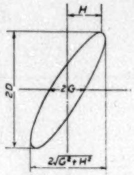


Fig. 9.

The critical 'turnover' frequency could be estimated fairly well by inspection of the oscillograph traces. It was found more accurately from the analysis of each photograph from which was obtained the resistive and inductive components at each frequency. This was done by measuring the lengths D, G and $(G^2+H^2)^{1/2}$ as indicated in fig. 9. It can then be shown that if the impedance of the point contact is R_c+iX_c then

$$\frac{R_c}{R}=\frac{H}{D}S\left(1+\frac{R_T}{R}\right)-1 \quad \text{and} \quad \frac{X_c}{R}=\frac{G}{D}S\left(1+\frac{R_T}{R}\right)$$

where R is the resistance in series with the point contact and S is the slope of the oscillograph line obtained when the contact is replaced by a resistance R_T.

As a check that the peak conditions of the d.c. characteristic are an indication of the true contact radius, a series of measurements was made with increasing mechanical pressure on the contact wire. Several experimental difficulties were encountered—firstly how to damp out the high frequency vibrations of the building, secondly how to avoid buckling of the contact wire and thirdly how to apply loads in the range 0·01 to 10 g gradually without 'impulse' and also to keep the 'zero load error' to a minimum. The first was overcome by mounting the apparatus on soft sponge rubber attached directly to one of the main girders of the building and the second by keeping the contact wire as short and straight as possible. The third difficulty was overcome by using a variable weight of liquid as the mechanical load. The apparatus was arranged with the contact wire vertical, a small open reservoir containing water rested on the block in which the upper end of the wire was gripped, and its weight was opposed by a light spring. A siphon tube connected this reservoir with a much larger fixed reservoir mounted nearby. The semiconductor disc was not allowed to touch the contact wire until readings were about to be taken; it was then slowly raised by means of a fine screw thread until it just touched the wire as indicated by the passage of current. This was regarded as the condition of zero load. Addition of a given weight of water to the larger reservoir resulted in one-twentieth of this weight being added to the load on the wire; the range 0·05 to 2 g was covered in this way.

3.2. *Results*

Several magnesium titanate discs having activation energy near one of the chosen values, 0·60, 0·48 and 0·34 ev were made the subject of d.c. measurements. Contacts at two or three points on each disc were observed. A typical voltage–current curve for a disc of 0·60 ev activation energy is shown in fig. 11. The results obtained with the tungsten point alternately positive and negative are rather unexpected. At low voltages and currents the characteristics are identical in the two directions, except with discs of the higher activation energy, when occasionally there is a very slight difference, the easy direction of flow being with tungsten positive. It can be seen that the relationship between current and voltage is not linear even at the lowest currents measured, namely 0·01 to 1 μA, but the current is always increasing more rapidly than the voltage. (The d.c. theory indicates that Ohm's law should be obeyed up to approximately

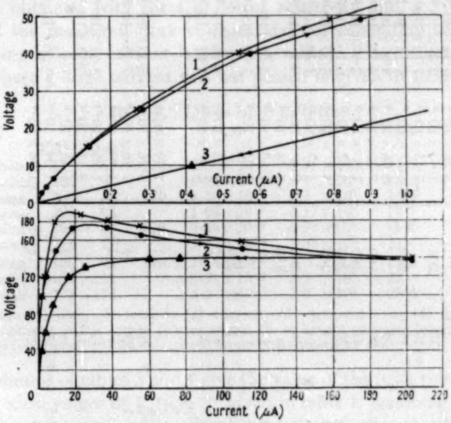


Fig. 11. Experimental voltage–current curve for a tungsten point to reduced Mg_2TiO_4(E=0·60 ev). Curve 1, increasing, tungsten positive; curve 2, increasing, tungsten negative; curve 3, decreasing.

Table 3

(1)		(2)		(3)		(4)	(5)		(6)	(7)		(8)	(9)
(*a*)	(*b*)	V_p(v)	I_p(μA)	V_p(v)	I_p(μA)	(μA)	W+ve	W−ve		W+ve	W−ve		
E~0·60 ev													
24	a	325	2	280	3·5	45	$2{\cdot}73\times10^{-6}$	$4{\cdot}12\times10^{-6}$	$1{\cdot}8\times10^{9}$	11	23	$1{\cdot}3\times10^{4}$	285
return		160	5	152	5·5		$3{\cdot}26\times10^{-6}$	$3{\cdot}52\times10^{-6}$	$3{\cdot}88\times10^{9}$	120	140		
24	b	190	18	176	30	206	$1{\cdot}44\times10^{-5}$	$2{\cdot}2\times10^{-5}$	$1{\cdot}6\times10^{8}$	13·5	27	$1{\cdot}3\times10^{4}$	285
return		142	~120	142	~120		$\sim7{\cdot}15\times10^{-5}$	$\sim7{\cdot}15\times10^{-5}$	$2{\cdot}3\times10^{7}$	18	18		
24	c	204	2	178	15	240	$1{\cdot}71\times10^{-6}$	$1{\cdot}12\times10^{-5}$	$4{\cdot}2\times10^{8}$	4·1	35	$1{\cdot}3\times10^{4}$	285
return		148	75	148	75		$4{\cdot}66\times10^{-5}$	$4{\cdot}66\times10^{-5}$	$4{\cdot}2\times10^{7}$	21	21		
repeat		172	100	154	70	160	$7{\cdot}22\times10^{-5}$	$4{\cdot}53\times10^{-5}$	$4{\cdot}2\times10^{7}$	24·5	19		
24	e	178	24	141	45	240	$1{\cdot}8\times10^{-5}$	$2{\cdot}67\times10^{-5}$	$1{\cdot}4\times10^{8}$	19	45	$1{\cdot}3\times10^{4}$	285
repeat		130	102	121	120	240	$5{\cdot}57\times10^{-5}$	$6{\cdot}1\times10^{-5}$	$3{\cdot}3\times10^{7}$	26	33		
24	f	370	3	320	4	240	$4{\cdot}7\times10^{-6}$	$5{\cdot}37\times10^{-6}$	$2{\cdot}43\times10^{9}$	20	30	$1{\cdot}3\times10^{4}$	285
return		270	13	254	14		$1{\cdot}47\times10^{-5}$	$1{\cdot}49\times10^{-5}$	$1{\cdot}22\times10^{8}$	5·9	6·7		
24	j	146	140	146	150	800	$8{\cdot}6\times10^{-5}$	$9{\cdot}2\times10^{-5}$	$1{\cdot}7\times10^{7}$	16	17·5	$1{\cdot}3\times10^{4}$	285
repeat		> 134	—	>134	—	850	$>4{\cdot}78\times10^{-4}$		$1{\cdot}54\times10^{6}$				
26	a	240	16	240	35	†	$1{\cdot}62\times10^{-5}$	$3{\cdot}53\times10^{-5}$	9×10^{7}	6	13	$8{\cdot}36\times10^{3}$	229
return		220	111	220	111		$1{\cdot}03\times10^{-4}$	$1{\cdot}03\times10^{-4}$	$1{\cdot}5\times10^{7}$	7·6	7·6		
repeat		220	111	220	111	200							
26	c	270	40	260	40	†	$4{\cdot}53\times10^{-5}$	$4{\cdot}36\times10^{-5}$	8×10^{7}	12	12·1	$8{\cdot}36\times10^{3}$	229
return		240	100	240	40		$1{\cdot}01\times10^{-4}$	$1{\cdot}01\times10^{-4}$	$1{\cdot}25\times10^{7}$	5·2	5·2		
E~0·48 ev													
195/1	4	> 50	—	—	—	1·1	$>1{\cdot}59\times10^{-4}$		9×10^{6}			1092	95
return		54	300	50·2	368		$4{\cdot}16\times10^{-5}$	$5{\cdot}32\times10^{-5}$	$1{\cdot}3\times10^{7}$	72	95		
repeat		51·1	500	51·1	500	†	$7{\cdot}36\times10^{-5}$	$7{\cdot}36\times10^{-5}$	$1{\cdot}3\times10^{7}$	127	127		
repeat		> 52·8	—	—	—	3·14	$>4{\cdot}79\times10^{-4}$		$3{\cdot}44\times10^{5}$				
195/1	6	58·3	116·4	56	112		$1{\cdot}96\times10^{-5}$	$1{\cdot}81\times10^{-5}$	$9{\cdot}5\times10^{6}$	19	19	1092	95
195/1	7	52	550	52·3	52·5		$8{\cdot}25\times10^{-5}$	$9{\cdot}91\times10^{-5}$	$8{\cdot}4\times10^{6}$	89	84	1092	95
195/1	8	> 45	—	—	—	3	$>4{\cdot}45\times10^{-4}$						
repeat		54·4	~6620	54·5	~10700	11·1	$\sim1{\cdot}04\times10^{-3}$	$\sim1{\cdot}68\times10^{-3}$	$3{\cdot}35\times10^{5}$	41	67	1092	95
repeat		> 52·6	—	—	—	6·95	$>1{\cdot}06\times10^{-3}$						
E~0·34 ev													
265/SO	1	43·5	~150	43·5	175	11	$\sim8{\cdot}87\times10^{-6}$	$\sim1{\cdot}04\times10^{-5}$	$2{\cdot}2\times10^{7}$	76	89	116·5	38·7
repeat		> 24	—	—	—	1·3	$>4{\cdot}25\times10^{-5}$						
265/SO	3	32·5	400	33·8	400	10	$\sim1{\cdot}77\times10^{-5}$	$1{\cdot}84\times10^{-5}$	$7{\cdot}1\times10^{6}$	87	84	116·5	38·7
repeat		> 43	—	—	—	10	$>1{\cdot}8\times10^{-4}$						
265/SO	4	> 25·6	—	—	—	1·3	$>4{\cdot}53\times10^{-5}$					116·5	38·7

(1) Identification (*a*) disc, (*b*) record; (2) tungsten +ve; (3) tungsten −ve; (4) maximum current passed, † a few mA; (5) contact radius (cm); (6) resistance R_0 **at lowest currents (ohms); (7)** R_0I_p/V_p; **(8) bulk resistivity (ohm cm); (9)** V_p **calculated from bulk resistivity (v).**

half the peak voltage.) As the current is further increased the two characteristics separate with tungsten negative always the easy direction of flow; there is also a marked time effect on reversing the current, several minutes being taken to establish equilibrium. The peak occurs at a higher voltage and smaller current in the tungsten positive direction, but the difference in voltage between the two directions becomes less marked as the current is still further increased.

When a current several times greater than the peak current has been passed through the contact, the peak is displaced to a lower voltage and much greater current and, at the same time, the difference between the tungsten negative and positive directions becomes less, and may even disappear—Ohm's law is then obeyed over the initial part of the characteristic. These effects can be simply explained if it is supposed that the contact radius becomes larger due to the flattening of the tungsten point at the higher temperatures.

If the highest current passed does not greatly exceed the current at the peak, there is no major difference between curves taken with ascending and with descending currents, but the voltage is usually a little higher when ascending.

Similar results, but with less difference between the characteristics in the two directions, are obtained with discs of lower resistivity and activation energy. A summary of the measured peak voltages and corresponding currents before and after passing a large current is given in table 3. A similar summary of experiments where a large current was not passed is given in table 4.

Table 4. Discs of $F \simeq 0{\cdot}6$ ev

(1)		(2)	(3)		(4)	(5)	(6)	(7)	(8)	(9)
(*a*)	(*b*)		V_p(v)	I_p(μA)						
79/1	a	Increasing	278	85	120	$9{\cdot}9 \times 10^{-5}$	$3{\cdot}75 \times 10^{7}$	11·5	$1{\cdot}33 \times 10^{4}$	288
		Decreasing	260	90		$9{\cdot}85 \times 10^{-5}$	$5{\cdot}3 \times 10^{7}$	18·3		
		Increasing	306	88	120	$1{\cdot}13 \times 10^{-4}$	$5{\cdot}3 \times 10^{7}$	15·4		
		Decreasing	293	88		$1{\cdot}08 \times 10^{-4}$	$6{\cdot}57 \times 10^{7}$	19·7		
		Increasing	286	82	120	$9{\cdot}86 \times 10^{-5}$	$6{\cdot}57 \times 10^{7}$	18·8		
		Decreasing	278	87		$1{\cdot}02 \times 10^{-4}$	$5{\cdot}85 \times 10^{7}$	21		
79/1	b	Increasing	180	95	130	$7{\cdot}18 \times 10^{-5}$	3×10^{7}	15·8	$1{\cdot}33 \times 10^{4}$	288
		Decreasing	169	120		$8{\cdot}5 \times 10^{-5}$	$2{\cdot}9 \times 10^{7}$	23		
		Increasing	175	173		$5{\cdot}4 \times 10^{-5}$	$2{\cdot}9 \times 10^{7}$	28·6		

(1) Identification (*a*) disc, (*b*) record; (2) current; (3) tungsten − ve; (4) maximum current passed (μA); (5) contact radius (cm); (6) resistance R_0 at lowest currents (ohms); (7) $R_0 I_p / V_p$; (8) bulk resistivity (ohm cm); (9) V_p calculated from bulk resistivity (v).

The last columns of tables 3 and 4 give the value of the peak voltage calculated from the theoretical values of $V_p(\rho_0\lambda)^{-1/2}$ given in table 1, assuming ρ_0 to be equal to the bulk value measured on the disc as a whole beforehand. The agreement with the measured values is not very good and it seems likely, from unpublished measurements of the impedance of similar discs over a wide range of frequencies, and from published measurements by Koops (1951) and Burgess (1952) on other sintered oxide semiconductors, that the material is composed of regions of various resistivities. It was therefore decided to regard ρ_0 as unknown, to be chosen so as to fit the experimental peak voltage. (This value of ρ_0 is of course a mean value, appropriate to the particular point of contact and small volume around it involved at the range of currents near to the value of the peak.) The ratio between this value and the bulk value mentioned above is equal to the square of the ratio between the measured and the calculated values of V_p in tables 3 and 4.

When scaled in this way the experimental results fitted the theoretical curves reasonably well as demonstrated in fig. 5, where the curves are those calculated in § **2.1**.

The value of the radius of the contact r_1 could now be calculated from the measured current at the peak and the theoretical value of x_1' at the peak, in fact it has been calculated from the product $V_p I_p$ using the value of K given in table 1 without the explicit calculation of ρ_0. The different position of the peak of the characteristic in the tungsten negative and positive directions, leads to two values of the contact radius as shown in table 3, but they are not widely different.

Alternating current ripple experiments were carried out on the discs of 0·60 ev activation energy appearing in tables 3 and 4, with the appropriate steady current, the tungsten being negative to the semiconductor (the behaviour was more stable with this polarity). Care was taken not to exceed twice the bias current, so that the d.c. peak was not changed before taking the a.c. ripple measurements. The peak voltages and corresponding currents are shown in table 5, together with the critical frequency at which the resistive component R_c of the impedance was zero.

Table 5

(1)		(2)	(3)	(4)	(5)	(6)
(*a*)	(*b*)					
24	1	395	87	10	$1{\cdot}45\times10^{-4}$	$1{\cdot}94\times10^{-4}$
24	2	361	55	42	$8{\cdot}34\times10^{-5}$	$9{\cdot}48\times10^{-5}$
24	3	536	60	15	$1{\cdot}35\times10^{-4}$	$1{\cdot}59\times10^{-4}$
26	2	108	35	1000	$1{\cdot}59\times10^{-5}$	$1{\cdot}94\times10^{-5}$
26	4	316	90	20	$1{\cdot}46\times10^{-4}$	$1{\cdot}37\times10^{-4}$
79/1	1	224	50	200	$4{\cdot}27\times10^{-5}$	$4{\cdot}35\times10^{-5}$
79/1	2	219	40	300	$3{\cdot}68\times10^{-5}$	$3{\cdot}55\times10^{-5}$

(1) Identification (*a*) disc, (*b*) record; (2) V_p (v); (3) I_p at V_p (μA); (4) critical frequency f_c (c/s); (5) contact radius from d.c. peak (cm); (6) contact radius from f_c (cm).

This critical frequency provides another value for ρ_0 or for r_1 using eqns. (18 and (19); the agreement with the ρ_0 or r_1 found from the d.c. peak indicate the degree to which the experimental conditions conform with theory. In table 5 the values of r_1 obtained in these two ways are shown; they agree within 25% or better.

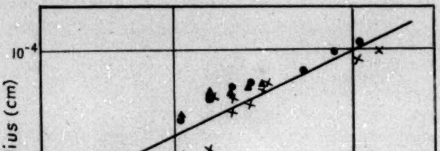


Fig. 12. Radius of platinum–rhodium contact against load.

Further, the values of A' of eqn. (18) and of $I_s R_c(\rho_0\lambda)^{-1/2}$ and $I_s X_c(\rho_0\lambda)^{-1/2}$ were calculated using ρ_0 as found from the d.c. peak voltage, and the points so obtained from a number of contacts are shown with the theoretical curve in fig. 7.

The measurements of the d.c. characteristic with various pressures on the contact point were also carried out with the metal negative to the semiconductor, and with the current limited to twice the peak current.

Using a tungsten point, there was in general no systematic variation in the peak conditions with increasing load. Microscopic examination of the metal point showed that the wire was breaking up near the contact region. In one case only, the radius increases as $(\text{load})^{0\cdot4}$ up to a load of 0·32 g corresponding to a stress of $2\cdot2 \times 10^6$ g cm^{-2}. At this point there was an apparent decrease in the contact area and the wire tip had broken.

Similar experiments were then performed using 0·015 cm diameter platinum–rhodium (10% Rh) wire. Systematic variation of the d.c. characteristic with load was now found, as long as care was taken to ensure the wire was straight. The results of measurements on three different wire points are shown in fig. 12, where it can be seen that the radius varies approximately as $(\text{load})^{0\cdot5}$ as expected if plastic flow occurs.

§4. DISCUSSION OF RESULTS

The experimental conditions evidently conform moderately well with the assumptions made in the theory of heating at the contact region, provided a value of resistivity to suit the conditions occurring near the peak of the voltage–current curve is assumed. This value is not the same as the value found from conventional measurements on the bulk material: it is sometimes larger and sometimes smaller but always of the same order within a factor of 3 or 4, despite the variation of 100 : 1 between the various specimens. The variation may be attributed to local variations of resistivity, for which there is other evidence.

The measurements of resistance R_0 at low voltages and currents are not, as a rule, consistent with the higher voltage results; R_0 is usually too high by factors up to 40, as shown by comparing $R_0 I_p / V_p$ in tables 3 and 4 with the calculated values in table 1. Moreover the voltage–current plot is usually curved and occasionally rectifying in this region. Mg_2TiO_4 is an n-type semiconductor and the direction of rectification is the one to be expected from a space-charge barrier at a metal contact. There is evidence that the density of donor centres is high in the range of resistivities used so that any space-charge barrier will be quite thin and its effect imperceptible when the total circuit voltage is high. It seems reasonable to suppose that the low voltage behaviour is due to space-charge barriers whose effects are not otherwise noticeable. It may be noted that the value of $R_0 I_p / V_p$ never falls below the theoretical value, and when it approaches it the calculated and experimental values of peak voltage are also in fair agreement.

The anomalous behaviour at high voltages is more surprising: rectification occurs in the opposite direction from that to be expected on space-charge barrier theories, and tends to disappear when the contact radius has apparently been enlarged by flattening at a somewhat higher temperature. (The effect of heating on the semiconductor itself would normally be to increase its resistivity.) Much of this could be explained in terms of field emission around the contact area.

Appreciable emission currents begin to flow whenever the surface field strength reaches about 10^7 v cm^{-1}. That such field strengths are possible over an appreciable area round the region of the contact can be shown by considering the end of the wire as part of a sphere of radius 10^{-3} cm (i.e. one-tenth of the diameter of the wire used), in contact with a plane surface over a circle of radius 10^{-4} cm. The gap between the spherical and plane surfaces would be less than 10^{-5} cm over a circle of radius considerably larger than the contact circle, so that with 200 volts or so applied to the contact the total current would be appreciably augmented. The exact calculation of the current flowing would be a matter of considerable difficulty because the surface of the semiconductor around the contact cannot be regarded as an equipotential.

Initially the emission current from a semiconductor will be greater than that from a metal since, in general, the work function of a semiconductor is less than that of a metal, but as there are fewer electrons available in the conduction band of the semiconductor compared with the metal, the emission current from the metal will ultimately exceed that from the semiconductor. The exact position of the crossover is very dependent on the values chosen for the various constants, particularly the work functions, which cannot be accurately known since they will depend on gases and vapours adsorbed on the surfaces. It seems likely that the crossover point will occur between 10^{-7} and 10^{-2} A cm^{-2} for a semiconductor having an activation energy of 0·60 ev. As the current density at the peak for these discs is of the order of 10^4 A cm^{-2}, emission from the metal would be greater than that from the semiconductor even at the lowest currents measured—that is, metal negative would be the easy direction of flow. This state of affairs has been discussed by Dilworth (1948) with the simplifying assumption that both surfaces are equipotentials.

The non-linearity of the voltage–current curve at voltages up to half the peak value is presumably due to the fact that the area over which emission is taking place increases as the voltage across the contact is increasing. At higher voltages the field emission component of the current from the metal when it is negative would lead to increased heating at the surface of the semiconductor by impinging electrons—the heating effect with the emission current in the other direction would be much less marked. The radius of the contact calculated from the peak voltage and current (and also from the a.c. ripple measurements) is therefore a measure of the effective area, including not only the metal–semiconductor junction but also the region of field emission. This may contribute to the discrepancy between measured and calculated values of $R_0 I_p / V_p$.

The slight hysteresis effect may be associated with field emission, since there is always less difference between the tungsten positive and negative directions when the current is decreasing. Remembering that the effective work function governing field emission will be that of a surface contaminated by adsorbed or combined gases, it may be that the heating produces a change of surface conditions such as to increase the current at a given voltage. This change may also be responsible for the time taken to establish equilibrium when the current is reversed. The adsorption or oxidation processes are themselves likely to display hysteresis.

When a current of several times the value at the peak voltage has been passed through the contact, there appears to be a considerable increase of contact radius and a diminution of ρ_0 as estimated from the peak voltage and current. It may be tentatively assumed that the former is due to the settling of the tungsten point on to the surface at the higher temperatures and that the latter is due to diffusion of oxygen from the Mg_2TiO_4 to the tungsten where it becomes adsorbed, but this point would need more investigation. The increase of radius and decrease of resistivity would lessen the importance of field emission round the contact area and tend to decrease the difference between the characteristics in the metal negative and positive directions at higher voltages, in agreement with observed facts.

ACKNOWLEDGMENTS

The author expresses thanks to Dr. R. W. Sillars who suggested this problem for many stimulating discussions, to her husband, Mr. D. P. Jones, for the preparation and bulk measurements of the semiconductors, to Miss B. M. Dent who suggested the methods of solution of the equations and for the computation involved. Thanks are also expressed to Dr. C. Dannatt, Director of Research and Education, and Mr. B. G. Churcher, Manager of the Research Department, Metropolitan-Vickers Electrical Co. Ltd., for permission to publish this paper.

REFERENCES

AMREIN, W., 1942, *Schweiz. Arch. angew. Wiss. Tech.*, **8,** 109, 152.
BENNETT, A. I., and HUNTER, L. P., 1950, *Phys. Rev.*, **81,** 152.
BENZER, S., 1949, *J. Appl. Phys.*, **20,** 804.
BURGESS, R. E., 1952, *Phys. Rev.*, **86,** 131.
DILWORTH, C. C., 1948, *Proc. Phys. Soc.*, **60,** 315.
KOOPS, C. G., 1951, *Phys. Rev.*, **83,** 121.
MITCHELL, E. W. J., 1952, *Proc. Phys. Soc.* B, **65,** 154.

Injected Absorption in Germanium

By A. F. GIBSON
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 19th January 1953

Abstract. When minority carriers are injected into germanium the density of current carriers is increased. This forms the basis of transistor action. The increased number of carriers is expected to cause increased absorption in the infra-red region of the spectrum. This effect has been observed and studied as a function of wavelength, injected current, frequency and other relevant parameters. The basic theory of the effect is discussed and is found to be substantially in agreement with the experimental results. Possible practical applications of the effect are described.

---

§ 1. INTRODUCTION

WHEN an electric field is applied to an ohmic conductor and a current flows, the total density of current carriers (electrons or holes, or both) remains unaltered and it is only the carrier drift velocity which is affected by the electric field. For semiconductors showing transistor action (e.g. Ge, Si, PbS), however, this is not so. The forward current of a crystal diode consists at least in part of minority carriers (for example positive holes) which enter the crystal in addition to the majority carriers (electrons) already present (Bardeen and Brattain 1949). Because of the space charge of the added number of minority carriers an equal number of majority carriers must be added from the base. Hence the total number of carriers is increased. In this respect transistor materials are unique.

The absorption spectra of several materials of this type have been measured in the infra-red by Becker and Fan (1949, 1950, 1951) (Ge and Si) and Gibson (1952) (PbS, PbSe and PbTe). In all cases the absorption spectra are characterized by a very sharp absorption edge at the limit of photoconductivity followed by a region of low absorption in which the absorption coefficient increases approximately as the square of the wavelength. Becker and Fan have proposed that this absorption is due to free carriers within the crystal and Gibson has applied the same interpretation to the PbS series. This theory predicts about the right wavelength dependence for the absorption but is, unfortunately, considerably in error as regards absolute magnitudes. Thus the absorption coefficient for unit density of carriers in germanium is about 1000 times greater than that predicted theoretically. This suggests that the carriers are not, in fact, completely free. Only in the case of the PbS series is there sufficient experimental data on the temperature dependence of the absorption coefficient for comparison with theory, but in this respect also the theory is inadequate.

According to Becker and Fan the theoretical absorption coefficient due to free carriers is

$$k=\lambda^2\left[\frac{\mathbf{e}^3N(\mu_0/k_0)^{1/2}}{4\pi^2\boldsymbol{n}\mathbf{c}^2\boldsymbol{m}^2\mu}\right] \qquad \ldots\ldots(1\,a)$$

in rationalized MKS units, where λ is the wavelength, $\mathbf{e}$ the electronic charge, N the carrier density, μ_0 and k_0 the permeability and permittivity of free space, $\boldsymbol{n}$ the refractive index, $\mathbf{c}$ the velocity of light, $\boldsymbol{m}$ the mass of a carrier and μ its mobility. Hence

$$k/N\lambda^2=\alpha/\lambda^2\simeq 0{\cdot}8\times 10^{-12} \qquad \ldots\ldots(1\,b)$$

for holes in germanium, where α is the absorption coefficient per unit carrier density. If $\lambda=2{\cdot}5$ microns, then from eqn. (1 *b*) $\alpha\simeq 5\times 10^{-24}$ m^2.

In view of the inadequate agreement between the simple theory and experiment in this field, particularly as regards the absolute magnitude of the absorption coefficient, it would seem desirable to check that the infra-red absorption is due to the current carriers by some method other than that employed hitherto. The theory of the absorption coefficient, together with the theory of the transistor, would suggest that it should be possible to modify the infra-red absorption of n-type germanium by carrier injection. This has been attempted and the effect observed. To observe injected absorption in germanium is not difficult. A specimen of n-type single crystal germanium is required having resistivity of not less than about 0·01 ohm metre and a minority carrier lifetime of not less than about 5 microseconds.

A point contact diode made with such material is illuminated with a steady beam of light, focused to a spot near the contact, and the light transmitted through the germanium (i.e. wavelengths greater than the absorption edge in germanium, namely $1{\cdot}8\,\mu$) is detected by a PbS photoconductive cell. If now a suitable a.c. voltage from a beat-frequency oscillator is applied to the germanium diode an audio signal from the PbS cell can be displayed directly on a cathode-ray oscillograph with sufficiently high gain amplifiers. In practice a tuned amplifier at 800 c/s having a bandwidth of about 50 c/s has generally been used, when a signal to noise ratio of 1000 : 1 can readily be achieved.

§ 2. THEORETICAL TREATMENT

Following the simple observation of injected absorption, which in itself provides evidence that at least some of the absorption at long wavelengths in germanium is due to current carriers, progress can only be made with the aid of a relatively simple mathematical analysis of the arrangement being used. As the germanium crystals generally employed have been cut into rectangular blocks there are only two arrangements which are of importance, namely (i) the light and the injected current enter and leave the crystal in the same direction x, or (ii) the light and the injected current enter and leave the crystal at right angles to one another (light in direction x, current in directions y or z). These two cases will now be considered.

(i) *Light and Current in Same Direction x*

It is assumed that the current density across the surface of the crystal yz is uniform. The density of injected carriers falls off in the x direction as $\exp(-t/\tau)$, where τ is the carrier lifetime and t is the time taken for the injected current carriers to reach any distance x into the crystal. Let the injected current be i and the specimen have dimensions x_0 (length) and A (cross section). Let R_0 be the resistivity and μ the mobility. It is assumed that all the forward current in the germanium diode is carried by minority carriers. Experiments by Dr. C. A. Hogarth of this laboratory show that for germanium crystals which have undergone similar treatment to those used, this assumption is substantially correct.

Then if the reduction in R_0 by the injected current is not large, the field produced by the current in the specimen is given by $F = iR_0/A$ and the time t taken for an injected carrier to reach a distance x into the crystal is therefore $xA/\mu iR_0$. Let t_0 be the value of t when $x = x_0$. The density of injected carriers then falls off in the direction x as $\exp(-xA/\mu iR_0\tau)$. Hence, if the *actual* density of carriers is known at, say, $x = 0$ then the density is known at any point x.

Let the density of injected carriers at $x = 0$ be P. To find P, let n be the total number of injected carriers in the rod in equilibrium and q the electronic charge. Then:

The number of holes entering the crystal per second $= i/q$.

The number of holes leaving the crystal per second $= (i/q)\exp(-t_0/\tau)$ and the rate of recombination of holes $= n/\tau$.

Therefore in equilibrium $n = \dfrac{i\tau}{q}\{1 - \exp(-t_0/\tau)\}$.

In order to obtain a simpler final solution we shall write

$$n = \frac{i\tau}{q}\left(\frac{t_0}{\tau + t_0}\right),$$

which is the same as the above expression for $\tau \ll t_0$ and $\tau \gg t_0$ and differs only by a factor 1·26 when $\tau = t_0$. But

$$n = \int_0^{x_0} AP \exp(-Ax/\mu iR_0\tau)\, dx$$

so that

$$\frac{i\tau t_0}{q(\tau + t_0)} = P\mu iR_0\tau[1 - \exp(-Ax_0/\mu iR_0\tau)],$$

which determines P. Therefore the density of injected carriers Nx in an element dx at a distance x into the crystal is given by

$$Nx = \frac{t_0 \exp(-Ax/\mu iR_0\tau)}{q\mu R_0(\tau + t_0)[1 - \exp(-Ax_0/\mu iR_0\tau)]}. \quad \ldots\ldots(2)$$

Ignoring for the moment the absorption due to free carriers present in the rod in the absence of injected carriers then if J_0 is the initial intensity and J_1 the transmitted intensity,

$$J_1/J_0 = \exp\left(-\int_0^{x_0} k\, dx\right) \quad \ldots\ldots(3)$$

where $k = \alpha Nx$ and α is defined in eqn. (1 *b*).

Equation (3) is the fundamental equation of this arrangement and as will be seen later it is at this point that the theories of case (i) and case (ii) diverge. We have from eqn. (2)

$$\int_0^{x_0} k\, dx = \frac{\alpha\tau x_0}{q\mu R_0(\tau + t_0)}.$$

Thus the effective absorption coefficient k_i due to the added carriers is given by

$$k_i = \frac{\alpha}{q\mu R_0}\left(\frac{\tau}{\tau + t_0}\right). \quad \ldots\ldots(4)$$

Now let J be the transmitted intensity in the absence of injected carriers and J_i be the transmitted intensity in the presence of injected carriers, the 'background' carriers now being included. Then

$$\frac{J}{J_0}=\exp\left(\frac{-\alpha x_0}{\mu q R_0}\right) \quad \text{and} \quad \frac{J_i}{J_0}=\exp\left[\frac{-\alpha x_0}{\mu q R_0}\left(1+\frac{\tau}{\tau+t_0}\right)\right].$$

We may now define two useful quantities which can be measured, namely, the signal per unit intensity S given by

$$S=\frac{J-J_i}{J_0}=\exp\left(\frac{-\alpha x_0}{\mu q R_0}\right)\left\{1-\exp\left[\frac{-\alpha x_0}{\mu q R_0}\left(\frac{\tau}{\tau+t_0}\right)\right]\right\} \quad \ldots\ldots(5)$$

and the efficiency ϵ given by

$$\epsilon=\frac{J-J_i}{J_0}\bigg/\frac{J}{J_0}=1-\exp\left[\frac{-\alpha x_0}{q\mu R_0}\left(\frac{\tau}{\tau+t_0}\right)\right]. \quad \ldots\ldots(6)$$

The latter is a more readily measurable quantity, not requiring a knowledge of the absolute values of J_0, etc. It will be noticed that eqns. (5) and (6) may be simplified by writing $1/q\mu R_0=n_0$, where n_0 is the density of carriers in the crystal without injection.

In both the above formulae the only term involving current is t_0, which is given by $t_0=x_0 A/\mu i R_0$.

Hence the signal and the efficiency do not increase indefinitely with current but only until $t_0 \ll \tau$. This is in fact observed. Physically this arises because the field created by the injected current also sweeps out the injected carriers. The condition $t_0<\tau$ can be achieved experimentally if τ is not less than about 50 μsec: for material of lower lifetime severe heating is encountered before a sufficiently large current can be achieved.

To complete the analysis of this case it can be assumed that the condition $t_0<\tau$ has been achieved, or $\tau/(\tau+t_0)=1$. It can then be shown that the signal (eqn. (5)) has a maximum value when

$$\alpha x_0/q\mu R_0=0{\cdot}7 \quad \ldots\ldots(7)$$

and the efficiency, which approaches 1 as $\alpha x_0/q\mu R_0$ increases, has a value of 0·5 (50%) when the signal is a maximum. Physically the signal maximum arises because increasing x_0 indefinitely leads to large absorption by the 'background' carriers. Care should be taken, however, when using eqn. (7) as the condition $t_0<\tau$ implies large injection currents and hence a significant reduction in R_0. Thus the assumptions upon which eqn. (7) is based are not valid under these conditions. In addition it is assumed that all the current is carried by holes, an assumption which is certainly not valid at sufficiently large injection currents. The current at which the electron contribution becomes important is a function of the surface barrier conditions and not readily calculable.

(ii) *Light and Current at Right Angles*

If the current enters the crystal in the direction x and the light in the direction y then the density of injected carriers in the element x is as already given. Ignoring, as before, the 'background' carriers, the added absorption due to the injected carriers is given by

$$\int_0^{x_0}\frac{J_i}{J_0}\,\frac{dx}{x_0}=\int_0^{x_0}\frac{\exp(-k)}{x_0}\,dx \quad \ldots\ldots(8)$$

which may be compared with eqn. (3). Unfortunately k in eqn. (8) involves $\exp x$ and the integral therefore cannot be expressed as a simple function. It can be seen however that when the condition $t_0 < \tau$ is achieved the density of injected carriers is virtually uniform within the rod and both arrangements tend to become identical. When this condition cannot be achieved (as is sometimes the case) arrangement (*b*) is preferred for practical reasons and arrangement (*a*) when theoretical predictions are to be checked.

§ 3. THE VARIATION IN EFFICIENCY WITH DRIVING CURRENT

The modulation efficiency ϵ as defined by eqn. (6) can be readily measured experimentally with the arrangement shown in fig. 1. If a PbS cell detector and glass optical system are employed the effective wavelength region is about 1·8 to 3·0 μ. These limits are set by the germanium and the PbS cell respectively. According to the results of Becker and Fan, already referred to, the value of α is largely independent of wavelength in this wavelength region. The efficiency of the germanium infra-red modulator is obtained by measuring the magnitude of the signal from the cell with the modulator being driven and dividing by the magnitude of the signal obtained when the light is interrupted mechanically at the same frequency, the germanium modulator being switched off but left in position. This measurement is independent of the light intensity and of cell and amplifier sensitivities.

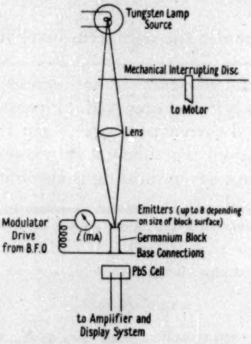

Tungsten Lamp Source
Mechanical Interrupting Disc
to Motor
Lens
Emitters (up to 8 depending on size of block surface)
Modulator Drive from B.F.O
i (mA)
Germanium Block
Base Connections
PbS Cell
to Amplifier and Display System

Fig. 1.

In arrangement (i), eqn. (6), two regions of current may be distinguished, namely currents such that $t_0 > \tau$ and $t_0 \simeq \tau$. The equation is not valid for $t_0 < \tau$. In the first low current region eqn. (6) reduces to $\epsilon = \text{efficiency} = i\alpha\tau/Aq$. As A and τ may be readily measured α can be determined as the slope of a straight line relating ϵ and i. Such a line, for a crystal of 0·07 ohm metre resistivity and 50 μsec lifetime, is shown in fig. 2. The value of α for this material is apparently $\alpha = 1{\cdot}5 \times 10^{-20}$ m^2. In comparing this value with the theoretical value of α for holes in germanium it must be remembered that for every hole injected an extra electron is also added. As electrons in germanium have about twice the mobility of holes, each added electron is, from eqn. (1 *a*), only half as effective as a hole.

Therefore the value of α for holes alone, deduced from the above experimental result, is $\alpha_{\text{holes}} \simeq \frac{2}{3}$ (α observed)$\simeq 1{\cdot}0 \times 10^{-20}$ m^2. It will be noticed that the straight line obtained in fig. 2 does not necessarily confirm the theory already given as probably almost any theory would indicate a linear relation between ϵ and i for sufficiently small drive currents. In fig. 3 however much larger drive currents (up to 100 mA) have been employed on the same specimen as before, and the full theoretical equation must be used. Equation (6) may be rearranged thus:

$$\frac{\alpha x_0 \tau}{\mu R_0 q}\left(\frac{1}{\tau + t_0}\right) = \ln\left(\frac{1}{1-\epsilon}\right)$$

and $1/(\tau + t_0)$ plotted against log $1/(1-\epsilon)$, as in fig. 3. The value of α so obtained is $1{\cdot}8 \times 10^{-20}$ m^2 or $\alpha = 1{\cdot}2 \times 10^{-20}$ m^2 for holes alone. This figure is in adequate agreement with that determined from fig. 2.

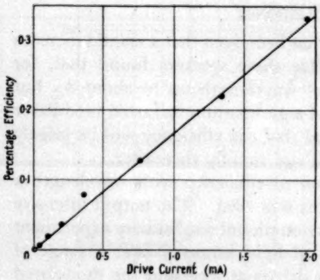


Fig. 2.

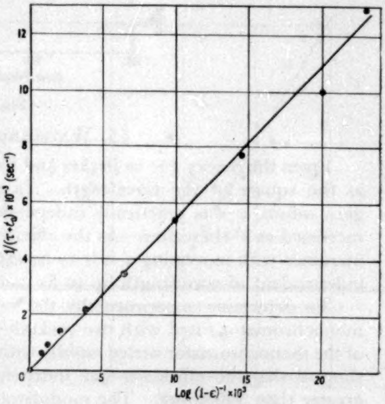


Fig. 3.

The theoretical value of α for holes in germanium in the 2–3 μ region from eqns. (1 *a*) and (1 *b*) is about 5×10^{-24} m^2, and hence the experimental value just quoted is about 2000 times larger than the theoretical value. For other samples the measured α has not been quite so large, but in all specimens there has been a factor of about 1000 between the simple theory and experiment. This is in agreement with the results of Becker and Fan already referred to and seems to confirm beyond reasonable doubt that the long wavelength absorption in germanium is mainly due to current carriers.

§4. FREQUENCY RESPONSE

The frequency response of an infra-red modulator relying upon injected absorption will be largely determined by the lifetime of the minority carriers. The lifetime will determine the rate at which injected carriers are removed during the half cycle in which there is no injection current. If the driving current is not too large the system is a linear one (as indicated by fig. 2) and the modulation depth will be reduced to a half at a limiting frequency given by $\jmath = 1/2\pi\tau$.

To determine the frequency response of a germanium modulator a sufficiently fast detector and wide band amplifier must be used. In view of the lack of a sensitive PbS cell with a time constant of less than about 20 μsec it has proved impossible to check the frequency response of really fast modulators. The frequency response of a relatively slow modulator is shown in fig. 4. It will be appreciated that high speed and the condition $t_0 < \tau$ are usually mutually exclusive. Nevertheless a bandwidth of 3 or 4 kc/s can be obtained without appreciable loss in efficiency.

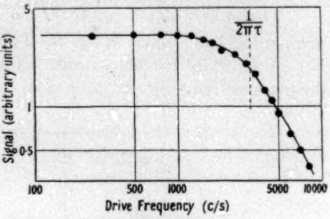


Fig. 4.

§5. WAVELENGTH RESPONSE

From the theory due to Becker and Fan it has been seen that α should increase as the square of the wavelength. In practice these workers found that, for germanium, α was practically independent of wavelength up to about 5 μ but increased as λ^2 thereafter. As the efficiency of a germanium infra-red modulator increases with increasing α it is to be expected that the efficiency will be largely independent of wavelength up to 5 μ and will rise rapidly thereafter.

To determine experimentally the variation in efficiency with wavelength a monochromator fitted with two rocksalt prisms was used. The output intensity of the monochromator varied rapidly with wavelength but a subsidiary experiment showed that the efficiency was independent of light intensity over a range of greater than 100 times. The modulator was driven at 5 c/s and the modulated light detected by a Schwarz thermopile followed by a tuned 5 c/s amplifier of the type designed by Brown (1949) of this establishment.

As the output intensity of the monochromator is rather low at the longest wavelength employed it is desirable for increased accuracy to use a modulator whose maximum signal (i.e. 50% efficiency) occurs at or about that wavelength. The efficiency of such a unit is shown in fig. 5. It will be seen that the qualitative predictions given above are largely substantiated. The two curves refer to different drive currents. As i, τ, R_0, etc., are known the value of α may be determined at each wavelength. Naturally both curves should give the same value of α at each wavelength. In fig. 6 the values of α so determined have been plotted against the square of the wavelength for wavelengths greater than 7 μ. The straight line obtained, which passes through the origin, substantiates the results of Becker and Fan. The slope of this line is, however, as follows:

$$\text{Experimental slope} = 1{\cdot}67 \times 10^{-9} \text{ or } 1{\cdot}1 \times 10^{-9} \text{ for holes alone.}$$

$$\begin{aligned}\text{Theoretical slope} &= \boldsymbol{e}^3(\mu_0/k_0)^{1/2}/4\pi^2 \boldsymbol{n}\mathbf{c}^2 m^2 \mu \\ &= 0{\cdot}8 \times 10^{-12}.\end{aligned}$$

As before, therefore, there is a discrepancy of about 1000 times, as would be expected from the previous measurements quoted.

§ 6. THE DESIGN OF A PRACTICAL MODULATOR

It has already been shown that, to obtain an efficient modulator, we must have (*a*) $t_0 < \tau$ and (*b*) $\alpha x_0/q\mu R_0 \simeq 0{\cdot}7$. To achieve the first condition τ must not be less than about 50 μsec unless very severe heating from the large driving current can be tolerated. The difficulty of achieving the second condition can be illustrated by means of an example. Suppose $\alpha = 1{\cdot}5 \times 10^{-20}$ m^2 and $R_0 = 5$ ohm cm. Then the optimum value of x_0 is given by $x_0 = 0{\cdot}7 q\mu R_0/\alpha = 6$ cm, which is an impracticable thickness. A reduction in the resistivity to, say, 1 ohm cm would bring the thickness to more practical values, but if the reduction in resistivity brought about a reduction in τ also, more might be lost than gained. It will be appreciated that large thicknesses of germanium are not required at long wavelengths when α is considerably larger than the value quoted above. Efficiencies of 50% can be readily achieved at 12 μ (fig. 5).

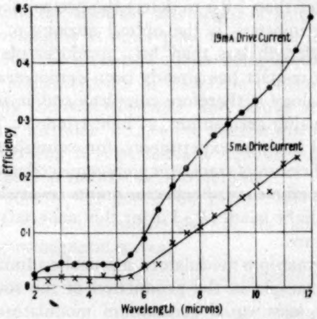


Fig. 5.

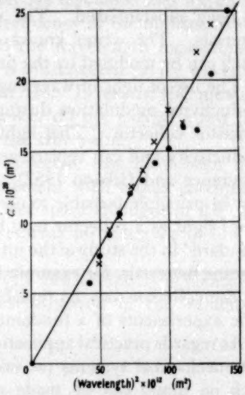


Fig. 6.

A small improvement in efficiency can be obtained by raising the temperature slightly above room temperature. This results in a reduction in μ and an increase in α, but an optimum is soon reached, probably due to a fall in emitter point efficiency, or lifetime, or both.

When practical infra-red modulators are to be made arrangement (ii), discussed before, is preferred. This arises as follows. When large crystal thicknesses are used to increase efficiency, t_0 is increased, thus making condition (*a*) above more difficult to achieve. By employing side injection however the distance through which the light has to travel can be kept large and the current flow distance small. Practical modulators, therefore, are made from long rods of germanium. Light, entering the rod at one end, is kept in the rod by total internal reflection at the walls of the rod. Because of the high refractive index of germanium quite large angles of incidence can be tolerated. In addition, side injection allows 'blooming' techniques to be employed to reduce reflection losses at the germanium crystal surfaces through which the light passes.

When operating germanium modulators in an infra-red telephone system it has been found desirable to apply about 2 volts d.c. bias to the modulators in the low resistance direction considered as a diode. If this is not done severe distortion can occur. An additional advantage of d.c. bias is to improve the frequency response of the modulator. This arises because the effective lifetime of the injected carriers is reduced to

$$\tau_{\text{eff}}=\frac{\tau t_0}{\tau+t_0} \quad \text{or} \quad f=\frac{\tau+t_0}{2\pi\tau t_0}$$

where t_0 is determined by the mean bias current.

§ 7. CONCLUSION

From the experiments carried out and the adequate agreement with the theory of injected absorption which has been outlined it is concluded that the optical absorption in the wavelength region from 1·8 μ to 12·0 μ in n-type germanium at room temperature is almost entirely due to free current carriers. The large numerical discrepancy in the theoretical equation (1 *a*) found by Becker and Fan has been substantiated. The experiments stress the uniqueness of transistor materials. The writer knows of no other materials the optical absorption of which can be modified by the passage of an electric current.

The use of light of wavelength greater than 1·8 μ to detect the occurrence of conductivity modulation during carrier injection is the optical analogy of the transistor collector. That light of wavelength less than 1·8 μ produces photoconductivity and can replace an emitter contact has already been demonstrated (Lawrance and Gibson 1952). The analogy is therefore complete and it is at least in principle possible to construct, using germanium, a 'light triode'. The use of light as a collector may be useful in some experiments, for example as a 'standard' in the study of the intrinsic efficiency of transistor point-type collectors. In some materials, for example PbS, the efficiency of collector points is very low, but the collector may be replaced by a light beam ($\lambda > 3\,\mu$ for this material) for some experiments of a fundamental nature.

As regards practical application of germanium modulators, it must be admitted that mechanical systems (as used for example in the production of the sound track on films) can be made which at least equal germanium modulators at relatively low frequencies. Germanium modulators have the advantages of small size, no moving parts and low power consumption (about 50 mw), but it is doubtful if these advantages outweigh the disadvantages in many practical cases. The wavelength limitation (beyond 1·8 μ only) is serious. Silicon would be better in this respect ($\lambda > 1{\cdot}2\,\mu$) and SiC better still ($\lambda > 0{\cdot}45\,\mu$), but these materials require further study.

ACKNOWLEDGMENTS

The author is indebted to his colleagues at the Telecommunication Research Establishment for help in some aspects of this work. He is indebted to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

BARDEEN, J., and BRATTAIN, W. H., 1949, *Phys. Rev.*, **75,** 1208.

BECKER, M., and FAN, H. Y., 1949, *Phys. Rev.*, **76,** 1530; 1950, *Ibid.*, **78,** 178; 1951, *Semiconducting Materials*, ed. H. K. Henisch (London: Butterworths Scientific Publications), p. 132.

BROWN, D. A. H., 1949, *J. Sci. Instrum.*, **26,** 194.

GIBSON, A. F., 1952, *Proc. Phys. Soc.* B, **65,** 378.

LAWRANCE, R., and GIBSON, A. F., 1952, *Proc. Phys. Soc.* B, **65,** 994.

The Fabry–Perot Interferometer at Millimetre Wavelengths

By W. CULSHAW
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 22*nd January* 1953, *and in amended form* 15*th April* 1953

Abstract. The design and operation of an interferometer of the Fabry–Perot type at wavelengths around 8 mm is discussed. The reflectors have been made using multiple quarter wavelength sheets of dielectric, and extremely sharp fringes have been obtained giving selectivity or Q factors exceeding 60000.

The effects of diffraction on the interferometer are discussed and wavelength measurements made at various positions. The frequency of the millimetre wave source has been stabilized to a high degree and the wavelength measurements at wider reflector spacings shown to agree with a substandard cavity wavemeter within the accuracy attempted of a few parts in 10^5. Errors in the measured wavelengths at other positions of the interferometer do not exceed a few parts in 10^4, the measured values being greater than the true value.

Measurements of dielectric constants and dielectric loss have been made using the interferometer, and results are in agreement with values obtained by other methods. Since the interferometer is the free space analogue of the cavity wavemeter it is particularly suitable for measurements on materials of low loss tangent at millimetre wavelengths.

Finally, the use of the interferometer for length measurements is mentioned, and also its possible use for the accurate determination of the velocity of electromagnetic waves.

§ 1. INTRODUCTION

In the microwave region, where the apertures of mirrors and lenses can readily be made somewhat greater than the wavelength, the use of microwave interferometers based on optical principles becomes feasible. The apertures which can be used in such microwave interferometers are, however, very much smaller in terms of the wavelength than those normally used in optical instruments, and a consideration of the diffraction which inevitably occurs is most important in their design and use.

The principles of the Michelson interferometer have been adapted for use at these wavelengths, and some measurements of wavelength and dielectric constants using such an instrument have been made (Lengyel 1949, Culshaw 1950). In general these show that the measured wavelength using the interferometer is too large, but approaches the correct value as the spacing of the instrument is increased. For path lengths of some metres this error amounts to a few parts in 10^4. More recently Froome (1952) has used both waveguide and optical principles in a Michelson interferometer to determine the velocity of electromagnetic waves at 1·25 cm. With aperture to reflector spacings up to 20 metres, together with a correction for diffraction effects, he has obtained excellent agreement with recent determinations by other methods.

It seems clear from these results that to obtain accurate wavelength measurements using microwave analogues of the Michelson interferometer requires relatively large spacings of the component parts of the instrument, and it is necessary to consider whether the application of other optical interferometers to these wavelengths can lead to a reduction in these space requirements, as well as to improved discrimination of the 'fringes'. With these aims in view the design and use at wavelengths around 8 millimetres of a Fabry–Perot interferometer is discussed. The 'fringes' of this interferometer are made much sharper by the use of multiple interference between two highly reflecting surfaces, and thus in the microwave region it forms the free space analogue of the resonant cavity. One of the problems in this application is the design of reflectors having the required high reflectivity without serious absorption of the microwave radiation. For this reason, as well as the size required, the use of thin uniform metal films does not appear promising.

One such interferometer in which the fringes are viewed by reflection is described by Sachs, Artman and Richter (1951). The reflectors consisted of metal balls embedded in polyfoam or silver squares on dielectric sheets. Such reflectors seem difficult to make, and the sensitivities obtained were not very high. Some measurements of wavelength were made with the interferometer but no estimate of their accuracy was given. In the present interferometer a solution to the reflector problem has been found by using multiple quarter wave sheets of dielectric spaced a quarter wavelength apart in air. Such a reflector enables very high reflectivities to be readily obtained, the limit being set by the dielectric loss of the material used. In this way, by using dielectrics such as fused quartz, reflectivities exceeding that of silver may be obtained without any serious absorption of the microwave radiation. The design of these reflectors will be discussed together with the effects of diffraction on the design and operation of the interferometer.

§2. THEORETICAL CONSIDERATIONS

(i) *Transmitted Intensity and Sharpness of Fringes*

The operation of this interferometer is due to multiple reflections between two surfaces or films. Consider incident plane waves, and let P and R be the transmission and reflection coefficients of intensity of the films; then the resultant intensity transmitted through both films is given by the Airy formula (Tolansky 1948)

$$T=\frac{P^2}{1-2R\cos\phi+R^2} \qquad \ldots\ldots(1)$$

where $\phi=(2\pi/\lambda)2\mu t\cos\theta$ is the phase difference between consecutive beams, t is the distance apart of the films, μ the refractive index of the medium between the plates, and θ the angle of incidence on the films. For maximum transmitted intensity $\cos\phi=1$ or

$$2\mu t\cos\theta=n\lambda \qquad \ldots\ldots(2)$$

where n is the order of interference, and

$$T_{\max}=P^2/(1-R)^2, \qquad \ldots\ldots(3)$$

which becomes unity for zero loss in the films.

Defining the Q value measured on the interferometer as the wavelength divided by the distance between points of half maximum intensity either side of resonance gives, using eqns. (1) and (2),

$$Q=\pi/\phi_1 \qquad \ldots\ldots(4)$$

where

$$\cos\phi_1=\{2R-(1-R)^2\}/2R \qquad \ldots\ldots(5)$$

The values of Q for various values of R are:

R	0·50	0·90	0·99	0·999
Q	4·5	30	432	2700

It is thus clear that to obtain sharp fringes reflectors of high reflectivity and little absorption are required.

(ii) *Development of Reflectors*

If the electric field in a plane wave travelling in the x direction is given by the real part of $E=E_0\exp(i\omega t-\gamma_x)$, where $\gamma=\alpha+i\beta$ is the propagation constant, then if $R_{n+1,\,n}$ be the amplitude reflection coefficient at the boundary between infinite media $n+1$ and n, and A_n be the amplitude reflection coefficient at the boundary between infinite media n and $n-1$ (see fig. 1), it can be shown by the method of multiple reflections, or by impedance considerations (Stratton 1941), that

$$A_{n+1}=\frac{R_{n+1,\,n}+\exp(-2i\gamma_n a_n)A_n}{1+R_{n+1,\,n}\exp(-2i\gamma_n a_n)A_n} \qquad \ldots\ldots(6)$$

where a_n is the thickness of the nth layer. This is the fundamental formula for the transfer of reflection coefficients through dielectric layers.

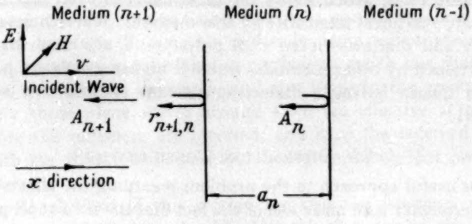


Fig. 1. Arrangement for calculation of reflection coefficients.

Note. $r_{n+1,\,n}$ should read $R_{n+1,\,n}$.

Applying eqn. (6) to determine the reflection from a quarter wave sheet of dielectric bounded on either side by free space, and changing the notation so that A_1 refers to one such sheet,

$$A_1=R_{01}(1+\psi)/(1+R_{01}{}^2\psi) \qquad \ldots\ldots(7)$$

where $\psi=\exp(-\pi\tan\delta/2)$, $\tan\delta$ is the loss tangent of the dielectric, and R_{01} the amplitude reflection coefficient at a free space–dielectric boundary. For $\tan\delta=0$, i.e. no dielectric loss, this reduces to

$$A_1=(1-\epsilon)/(1+\epsilon) \qquad \ldots\ldots(8)$$

where ϵ is the dielectric constant. Generally the amplitude reflection coefficient of n quarter wave dielectric sheets spaced quarter wave in free space apart is equal to that of a single quarter wave sheet having a dielectric constant of ϵ^n.

However, in view of the high reflectivities involved, the effects of a finite dielectric loss must be considered. The main effect of this will be to attenuate the waves passing through the sheets. Applying eqn. (6) to transform the amplitude reflection coefficient of n such quarter wave sheets through a quarter wavelength in free space followed by an additional quarter wave sheet gives the general formula used to consider the effect of dielectric loss:

$$A_{n+1}=\frac{A_1+A_n(R_{01}{}^2+\psi)/(1+R_{01}{}^2\psi)}{1+A_1A_n}. \qquad \ldots\ldots(9)$$

This tends to a limit as n becomes large given by

$$A=\frac{-(1-a)+\{(1-a)^2+4c^2\}^{1/2}}{2c} \qquad \ldots\ldots(10)$$

where $c=A_1$ and $a=(R_{01}{}^2+\psi)/(1+R_{01}{}^2\psi)$.

A consideration of polystyrene for which $\epsilon=2{\cdot}56$ and $\tan\delta=0{\cdot}001$ approximately at $\lambda=1{\cdot}25$ cm leads to the following values of the reflection coefficient A_n for numbers n of such quarter wave sheets, the limiting value being 0·9982:

n	1	2	3	4	5	6	7	8
A_n	0·4375	0·734	0·8861	0·9528	0·9806	0·9913	0·9961	0·9977

Thus eight quarter wave sheets of polystyrene each spaced a quarter wavelength apart in air will give an *intensity* reflection coefficient R of 0·9954 and a corresponding Q value from eqns. (4) and (5) of 675. If this dielectric had zero loss then the reflection coefficient of intensity from these sheets would be 0·9978 and the Q value 1350, which shows the great effect of even small loss tangents on the ultimate sharpness attainable by this method. Nevertheless, these values of reflectivity and sharpness, even with polystyrene, are much higher than any previously attained by other methods, and still higher values are possible by the use of fused quartz having a dielectric constant of 3·83 and a loss tangent around 0·0001.

§3. DIFFRACTION CONSIDERATIONS

The most useful approach to the problem regarding the effects of diffraction on the interferometer is to make use of the fact that the field at all points in front of a plane aperture of any field distribution may be regarded as arising from the interference of plane waves in various directions. The amplitude and phase of these waves, expressed as a function of their direction of travel, constitute an angular spectrum of radiation which, appropriately expressed, is the Fourier transform of the aperture distribution (Booker and Clemmow 1950).

If the aperture is of limited extent, as in aerial systems, then at distances from the aperture large compared with its width and to the wavelength the only plane wave effective is that in the direction of the distant point. As the distance from the aperture decreases other plane waves in the spectrum of the aperture distribution become effective and must be considered in determining the resultant field at a given point.

In the aperture illuminations envisaged for the interferometer the electric field is uniform in phase and amplitude in the vertical or E plane, whilst in the horizontal plane the amplitude varies sinusoidally, the phase again being uniform.

The effects of diffraction may thus be assessed by considering the number of plane waves which form the angular spectra of these aperture distributions, and what their effect is likely to be on the interferometer.

The theory developed in § 2 shows that the reflectors will exert a selective action on these plane waves of the angular spectra, since the path difference between adjacent beams in the interferometer varies as cos θ. Thus for high reflectivities only a small amount of the angular spectrum, depending also on the order of interference, is effective between the reflectors, the remaining portion being out of phase. It follows from eqns. (2) and (4) that the angular width of the spectrum transmitted between points of half maximum intensity is given by

$$\cos\theta = 1 - 1/2nQ. \qquad \ldots\ldots(11)$$

Thus the higher the Q value and order n of interference the greater this selection, the process being strictly analogous to what occurs in a resonant cavity which can propagate a number of modes. Due also to diffraction effects energy will be lost outside the reflector system, the amount lost increasing with reflector separation. The responses should therefore be less sharp or the Q value should decrease as the separation of the reflectors increases.

One further effect of diffraction on the interferometer will be to limit the fraction of intensity transmitted at a maximum. Although the effective conductivity of the dielectric used limits the reflectivity attainable by the method described in § 2 (ii), the actual absorption of energy by the overall thickness of dielectric used in the reflectors is quite small, and since at a maximum transmission the reflectivity or high impedance of one reflector is, at least at the design frequency, exactly annulled by the second, then for a single plane wave along the axis the transmission coefficient of intensity should approach unity. However, as pointed out, the radiated energy is spread out into the angular spectrum of the aperture illumination and only plane waves around $\theta = 0$ will be transmitted, resulting in transmission coefficients much less than unity. At large distances between the apertures, and also between the apertures and the reflectors, only those plane waves around $\theta = 0$ are effective at the receiving aperture before the reflectors are inserted, and thus the received energy at a maximum with the reflectors in position should approach that obtained when they are removed.

The main conclusions of this section are that the positions of maximum response on the interferometer should approach that due to the fundamental $\theta = 0$ plane wave, especially at the higher orders of interference, and hence accurate measurements of wavelength should be possible. There may be irregularities, however, if the order of interference becomes small so that plane waves at other angles become effective between the reflectors. The Q value should decrease with increasing reflector separation and the transmission coefficient of intensity will be much less than unity.

§4. DESIGN AND OPERATION OF INTERFEROMETER

(i) *Description of Interferometer*

A schematic diagram of the interferometer is shown in fig. 2, the fringes being viewed by transmission. The radiating and receiving apertures are made by tapering up rectangular waveguide containing the H_{10} mode to form an electromagnetic horn. The amplitude configuration at the aperture is thus essentially that of the H_{10} mode in the waveguide, i.e. uniform in the vertical or E plane through the guide axis and sinusoidal in the horizontal plane. The required uniform phase across the aperture plane is obtained with a dielectric lens in the aperture, of the appropriate focal length.

In this way apertures of dimensions $3\frac{1}{2}$ in. × $3\frac{1}{2}$ in. and 6 in. × 6 in. were provided so that the width of the angular spectrum radiated could be varied to investigate its effect on the measurements. The lenses for these apertures were made from polystyrene and were designed using geometrical optics, the focal lengths being 15 cm and 50 cm respectively. The surfaces of these lenses were 'bloomed' or matched to free space so as to eliminate reflections from them, and the aperture perimeter was bevelled at 45° to reflect stray energy outside the system. However, in view of the high reflectivities of the main reflectors, and the method of measurement, it is felt that the effect of such aperture reflections would in any event be small.

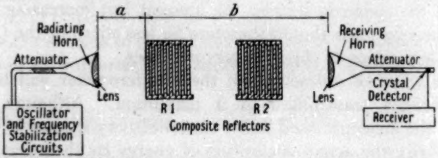


Fig. 2. Schematic diagram of millimetre wave Fabry–Perot interferometer.

A reflection klystron generating wavelengths around 8 millimetres was used as the source, and since accurate measurements of wavelength, in addition to the plotting out of sharp responses, were envisaged, the klystron was frequency stabilized to a high degree using a resonant cavity as a frequency discriminator after the manner described by Pound (1946). The received radiation was fed into a millimetre wave crystal detector and the rectified current observed on a galvanometer.

For these investigations the reflectors were made from quarter wave sheets of polystyrene made in disc form with a strengthening rim at the outside, which also provided the required quarter wave spacing of the sheets in air (fig. 2). The reflectors consisted of eight such sheets held together in a suitable frame with an effective diameter of 11 inches. The complete reflectors were mounted on a substantial V-slide 5 ft long, and levelling and slow motion adjustments were provided to bring the reflectors into the vertical plane at right angles to the axis of the slide. External supports for the horn radiators were also provided, and the smaller of these could also be mounted on the slide. One of the reflectors, R2, could be moved along the slide by means of a lead screw, and a reading of accuracy to 10^{-4} cm was provided by a dial and vernier. Locating holes at intervals of 50 cm were provided for the rapid movement of this reflector.

(ii) *Operation of Interferometer*

Since the interferometer operates with an essentially parallel beam, apart from the difficulties of diffraction, the circular fringe system of the optical model is not obtained, instead the single fringe is observed by altering the distance between the reflectors. The microwave radiation can be made highly monochromatic by frequency stabilization, and the individual fringe shapes are readily measured. The visibility of the fringes at large path differences is thus very good and fringes are readily observed over path lengths of some metres.

To set up the interferometer the components are placed in line with the slide, and orientation and levelling of the reflectors carried out until a fringe is obtained with the reflectors spaced some 50 cm apart. This adjustment, as can be expected with the high reflectivities involved, is rather critical, and is continued until the sharpest possible fringe is obtained.

In wavelength measurements using the interferometer the initial reflector separation was adjusted for a maximum using the lead screw, and then reflector R2 was moved quickly into the next locating hole 50 cm away, and adjusted on the nearest maximum by the lead screw. Both the movement of the lead screw and the distances apart of the locating holes were measured by length gauges to an accuracy of 1 in 10^5. Three locating holes with accurately known spacings around 50 cm were thus provided and measurements could be made rapidly and accurately by this means after an initial determination of the number of fringes in the intervals.

Measurements of the apparent wavelength in air on the interferometer were thus made moving R2 in both directions and means taken. These were repeated in each position with different initial reflector separations. Actual results were very consistent, due to the sharpness of the fringes and the frequency stabilization employed, setting accuracies to about $1\,\mu$ being readily achieved at most spacings.

The wavelength in air was also measured using an electroformed substandard H_{01} cavity wavemeter checked by reference to a crystal controlled oscillator and its accuracy verified to an order of a few parts in 10^5. This was considered adequate for these experiments, and thus a direct comparison of the two measured wavelengths could be made.

A measure of the dielectric constant and loss tangent of materials is obtained by inserting sheets between the reflectors and noting the shift of a given fringe and the decrease in sharpness. Some allowance, however, must be made for the fact that the shift of the fringe and also its sharpness will depend on the position of the sheet between the reflectors. This will be dealt with later in connection with the results obtained.

§ 5. RESULTS

(i) *Wavelength Measurements*

In order to investigate the effects of diffraction on the interferometer, measurements of the apparent wavelength were made with the radiating and receiving horns at the following distances, *a* and *b* respectively (fig. 2), from the first reflector R1.

Position	1	2	3	4	5
a (cm)	60	115	33	45	9·5
b (cm)	300	230	170	170	150

Figure 3 shows the sharpness of the fringes obtained in relation to the wavelength. These were measured using the 6 in. × 6 in. apertures at a reflector separation of 12 cm.

Measurements of the wavelength in air at positions 1 and 2 using these apertures for various initial reflector separations are as follows, the measurements below an initial separation of 74 cm being made at position 2. During these measurements the wavelength in air determined using the cavity wavemeter was 8·3299 mm. As was anticipated in § 3, there is thus close agreement at the larger initial reflector separations, or higher orders of interference, and it is seen that errors of a few parts in 10^4 can occur at smaller initial reflector separations, at least for the spacings of the radiating and receiving apertures employed here.

Initial reflector separation (cm)	8	16	31	66	74	124
Measured wavelength (mm) using 120 fringes	8·3334	8·3326	8·3324	8·3314	8·3306	8·3299

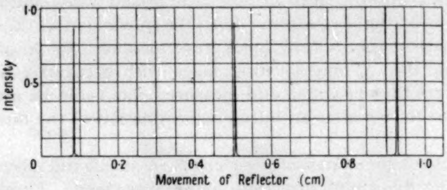


Fig. 3. Showing sharpness of fringes obtained on the interferometer in relation to the wavelength.

The actual responses obtained at reflector separations of 74 cm, 124 cm, and 174 cm are shown in fig. 4. It is seen that the responses obtained are quite sharp even at reflector separations around 2 metres, and that extremely high values of the selectivity factor in terms of a wavelength change, viz. $\lambda/d\lambda$ or $Qn/2$ from eqn. (2), are obtained. The accurate measurement of wavelength with the interferometer at these relatively small distances of the radiating and receiving horns is thus due to the high Q values and order of interference which have been obtained at wide reflector separations.

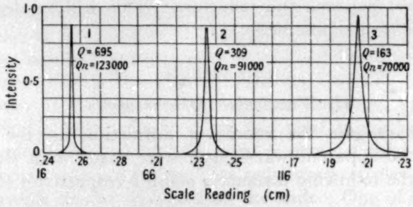


Fig. 4. Responses used at position 1 for wavelength measurements. Initial reflector separation 74 cm for curve 1.

Figure 5 shows responses every 10 fringes with the interferometer in position 3 and an initial reflector separation of 3 cm. The occurrence of the small secondary maxima and the increasing Q value at this spacing is to be noted. Referring to eqn. (2) for the path difference in the interferometer, the fact that these secondary maxima occur on the side of the response which corresponds to increasing reflector separation, and also that the distance between them and the main maxima increases with order, shows that these effects are presumably due to the increased effect of other plane waves of the angular spectrum at these small reflector spacings. The fall off in intensity of the fringes is shown in fig. 6. The intensity at first increases with reflector separation, in agreement with the observed increase in the Q value.

The effect of using radiating and receiving apertures of dimensions $3\frac{1}{2}$ in. × $3\frac{1}{2}$ in. was next investigated at position 4. From the responses Q values of 595, 362 and 177 respectively were deduced. These are slightly less than the previous values obtained using the 6 in. × 6 in. apertures at the similar position 3. The measured wavelengths were 8·3329 mm and 8·3319 mm for initial reflector separations of 18 cm and 68 cm respectively using as before 120 fringes in the measurements.

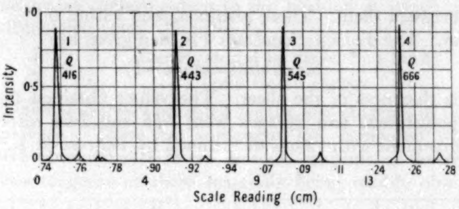


Fig. 5. Showing secondary maxima obtained with close reflector spacings. Position 3, 6 in. × 6 in. apertures.

These results are comparable with those obtained using the larger 6 in. × 6 in. apertures. The transfer of energy through the interferometer was, however, reduced when using these small apertures, since they have a wider angular spectrum; due to this reduction it was not possible with the present technique to measure the wavelength at larger spacings. The fall off in intensity of the fringes with reflector separation for the smaller apertures is shown in fig. 6 and shows a steeper descent than the similar curve for the larger apertures.

Finally measurements of wavelength were made using these smaller horns at position 5, with and without the phase correcting lenses at the horn apertures. These measurements were made using displacements of 120 fringes and initial reflector separations of 18 and 68 cm.

The results obtained were practically identical for the reflector spacings employed here. However, the energy transfer through the interferometer was reduced by some 60 db when the lenses were removed from the horn apertures, due no doubt to the much wider angular spectrum of the resulting aperture illumination. Its effect was clearly shown by an increase in both the amplitude and number of secondary maxima. The secondary maxima disappeared at a reflector separation of 43 cm.

(ii) *Measurement of Dielectric Properties*

The method of measurement of dielectric properties by means of the interferometer is analogous to the cavity resonator method normally used at centimetre wavelengths (Penrose 1946), and similar formulae apply, some simplification arising due to operating in free space. To measure the dielectric constant a sheet of the material is inserted between the reflectors and the shift of a given fringe noted. However, due to the differing impedances, effective path length changes occur at the boundaries of the sheet, the amount depending on the position of the sheet between the reflectors. The total shift of a fringe is thus periodic, and measurements must be made varying the position of the sheet by small intervals over a range of $\lambda/2$ and the mean fringe shift found. At the position of the mean shift the effective path length changes at the boundaries cancel out, and the true path length due to the thickness of the sheet is obtained.

The measurements were made in position 4 using the large apertures and 18 in. × 12 in. sheets of polystyrene and Perspex of thickness 0·865 in. and 0·825 in. respectively. The reflectors were spaced some 65 cm apart for the measurements, and the mean shift of a fringe due to the insertion of the polystyrene sheet was measured at three positions between the reflectors: near R1, near R2, and midway between them. These results agreed within 0·5%, the mean shift being 1·314 cm. The dielectric constant ϵ is then given by

$$\text{mean shift} = d(\sqrt{\epsilon} - 1) \qquad \ldots\ldots(12)$$

where d is the thickness of the sheet. The values obtained for the dielectric constants of polystyrene and Perspex were 2·55 and 2·64 respectively, which are in good agreement with values determined by Penrose at 1·25 cm.

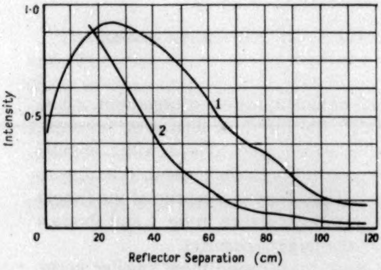


Fig. 6. Fall off in intensity of fringes with reflector separation. Curve 1, 6 in. × 6 in. apertures; curve 2, 3½ in. × 3½ in. apertures.

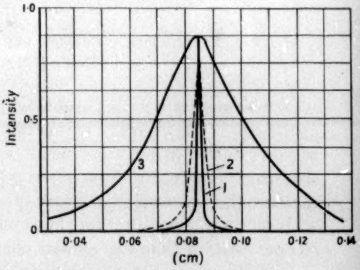


Fig. 7. Mean responses with and without dielectric sheets between reflectors. Curve 1, without dielectric; curve 2, polystyrene sheet; curve 3, Perspex sheet.

Similarly loss tangents are measured by regarding the interferometer as a free space resonator, and using formulae similar to those employed in the cavity resonator method. In such a resonator $1/Q_\lambda = B \tan\delta$ when resistive losses are neglected. Here Q_λ is the Q factor in terms of a wavelength change, given by $Q_\lambda = Qn/2$ for the interferometer (cf. eqn. (2)), and B is that fraction of the total stored energy present in the dielectric. This fraction depends both on the relative volume of dielectric present and the ratio A of field amplitudes in the two media, and as the position of the sheet between the reflectors varies the Q factor will change due to changes in the value of A. Assuming that one boundary of the sheet is at a node of the standing wave pattern, A is given by

$$A = [1 + (\epsilon - 1)\cos^2 2\pi d/\lambda_d]^{-1/2} \qquad \ldots\ldots(13)$$

where λ_d is the wavelength in the dielectric. Since the energy stored in the dielectric is proportional to A^2, the measured value of Q should vary periodically with position of the sheet by a factor of ϵ between maximum and minimum values.

Measurements were made using the polystyrene and Perspex sheets. An initial determination of Q without the sheets inserted gave a value Q_0 of 416, which is used to consider what might be termed the effective resistive loss in the interferometer. The order of interference n used was 158. The sheets were then inserted and the range of variation of Q measured. Figure 7 shows plots of the mean responses obtained with and without the sheets in position. The variations in the Q values were 124 to 273 for the polystyrene sheet and 11 to 35 for the Perspex sheet; in addition the transmission of energy through the interferometer was reduced by some 26 db on inserting the Perspex sheet. Taking the lowest Q values as corresponding to the position at which $A=1$, the loss tangents were calculated using the formula

$$\frac{1}{Q_\lambda}-\frac{1}{Q_{0\lambda}}=\frac{\text{Energy stored in dielectric}\times\tan\delta}{\text{Total stored energy}} \qquad \ldots\ldots(14)$$

and values of 0·0006 and 0·009 obtained for the loss tangents of polystyrene and Perspex respectively. These are somewhat low compared with the values 0·0008 and 0·012 obtained at 1·25 cm, but are certainly of the right order, the difference in loss tangents of these materials being readily observable on the interferometer.

§ 6. CONCLUSIONS

The design of reflectors using the multiple quarter wavelength technique has led to very high reflectivities in the interferometer, in agreement with theoretical predictions, and as a result extremely sharp fringes have been obtained with polystyrene which, though suitable for this initial work, is not the best material for the purpose. Expressed in terms of a wavelength change $\lambda/d\lambda$, selectivity or Q_λ factors exceeding 60000 have been obtained, which are comparable with those of the best cavity resonators at these wavelengths. It seems clear that by the use of a material with lower dielectric loss and higher dielectric constant, such as fused quartz, still higher reflectivities and sharper fringes can be obtained by this method. This would seem desirable for measurements at wider spacings of the interferometer than those which have been employed here.

Due to the reduction of diffraction effects, the wavelength measured on the interferometer at the wider reflector spacings agrees with that obtained using the substandard cavity wavemeter, within the accuracy attempted, to a few parts in 10^5. At close spacings of the interferometer errors of a few parts in 10^4 may arise due to the more pronounced effects of diffraction. These would not be serious for normal routine measurements involving the wavelength. In general, no discrepancy between measurements made using the larger and smaller apertures has been observed, but the larger apertures give both greater energy transfer and sharper fringes, and hence are desirable.

The measurements of dielectric constants are certainly within 1% of values obtained at these wavelengths by other methods, whilst the values of loss tangents obtained are in reasonable agreement with known values, and the method should be useful for measurements on materials of low loss tangent at millimetre wavelengths where conventional methods become difficult.

Since sharp fringes have been observed with the interferometer at distances of 2 metres between the reflectors, even with the polystyrene reflectors, it is felt that the interferometer possesses great potentialities for the accurate measurement of relatively large lengths—of some metres. It would also appear that with more robust and accurate reflectors of fused quartz this interferometer gives the best method for the accurate determination of the velocity of electromagnetic waves.

ACKNOWLEDGMENTS

The author would like to make the following acknowledgments: to Mr. C. R. Ditchfield who checked the substandard cavity wavemeter against a crystal-controlled oscillator, to Mr. S. Douglas who performed the length calibrations, and to Dr. R. A. Smith for his continued support and encouragement. Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller, H.M. Stationery Office, for permission to publish this paper.

REFERENCES

BOOKER, H. G., and CLEMMOW, P. C., 1950, *J. Instn. Elect. Engrs.*, Pt. III, **97,** 11.
CULSHAW, W., 1950, *Proc. Phys. Soc.* B, **63,** 939.
FROOME, K. D., 1952, *Proc. Roy. Soc.* A, **213,** 123.
LENGYEL, B. A., 1949, *Proc. Inst. Radio. Engrs., N.Y.*, **37,** 1242.
PENROSE, R. P., 1946, *Dielectrics* (London : The Faraday Society), pp. 108 *et seq.*
POUND, R. V., 1946, *Rev. Sci. Instrum.*, **17,** 490; *Proc. Inst. Radio Engrs.*, **35,** 1405.
SACHS, M., ARTMAN, J. O., and RICHTER, E., 1951, *Columbia University Radiation Laboratory Report,* June 1.
STRATTON, J. A., 1941, *Electromagnetic Theory* (London : McGraw-Hill), pp. 511 *et seq.*
TOLANSKY, S., 1948, *Multiple Beam Interferometry of Surfaces and Films* (Oxford : Clarendon Press), pp. 21 *et seq.*

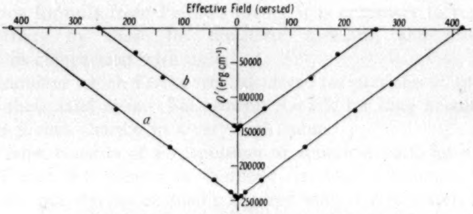
RESEARCH NOTES

Thermal Effects accompanying Magnetization of a Ferrimagnetic Material

By L. F. BATES and N. P. R. SHERRY
Physics Department, The University, Nottingham

MS. received 29*th April* 1953

Many measurements have been made on the thermal changes which occur when a ferromagnetic metal or alloy is taken step by step through an ordinary hysteresis cycle, but until recently samples of ferrimagnetic materials have not been available in the form required for such measurements. However, the G.E.C. Research Laboratories, Schenectady, very kindly gave us a rod of a mixed Ni–Zn ferrite, Ferroxcube IV (15% NiO, 35% ZnO, 50% Fe_2O_3), some 27 cm long and 0·6 cm in diameter, with which we have made satisfactory thermal measurements. The apparatus employed was that frequently used for investigating the behaviour of ferromagnetic materials (cf. Bates 1951, Bates and Marshall 1953), but in the present instance the apparatus was calibrated by heating the specimen by the passage of a small alternating current through it for a short time, as the more direct calibration by loading was not practicable.



Thermomagnetic measurements with Ni–Zn Ferrite.
Curve *a*, maximum field 410 oersteds.
Curve *b*, maximum field 50 oersteds.
(Note : all values for curve *b* have been multiplied by 5.)

The results reproduced in the figure show that although the overall heating for a half cycle is very small, substantial energy changes take place in all parts of the cycle. The fact that these large effects are a linear function of the effective magnetizing field clearly indicates that the changes are directly associated with changes in spontaneous magnetization. It is seen that even when observations are made in a cycle with a maximum field of only 50 oersteds, the results show no appreciable departure from the linear form, the slope of the graph being the same as in the case of larger cycles.

An approximate calculation based on the thermodynamic relation

$$(\partial Q/\partial H)_S = -T(\partial I_0/\partial T)$$

gives values in agreement with the measured thermal changes, if we assume that the Curie point is 104°c, that the value of the intrinsic magnetization at room temperature is 212 c.g.s. units and that it decays linearly to the Curie point (cf. Wijn 1953).

It is therefore concluded that although large thermal changes occur in this material, they are adequately explained in terms of the magneto-caloric effect of Weiss and Forrer and that, notwithstanding views expressed elsewhere (Wijn 1953), domain boundary phenomena do not appear to make a significant contribution to the energy changes, for the latter phenomena would give heating effects of sign opposite to those observed, and ought to show clearly on the graph.

REFERENCES

BATES, L. F., 1951, *J. Phys. Radium*, **12,** 459.

BATES, L. F., and MARSHALL, G., 1953, *Rev. Mod. Phys.*, **25,** 17. See also WIJN, H. P. J., GEVERS, M., and VAN DER BÜRGHT, C. M., 1953, *Rev. Mod. Phys.*, **25,** 91; RADO, G. T., *Ibid.*, 92; GALT, J. K., ANDRUS, J., and HOPPER, H. G., *Ibid.*, 93.

WIJN, H. P. J., 1953, *Dissertation*, University of Leiden, and Laboratoria N. V. Phillip's Gloeilampenfabrieken, Eindhoven; Separaat 2092.

LETTERS TO THE EDITOR

The Dielectric Constant of a Liquid containing Spherical Particles

In order to interpret microwave dielectric data on concentrated aqueous solutions of electrolytes, polar molecules (Haggis *et al.* 1952) and proteins (Buchanan *et al.* 1952), as well as such substances as blood (Cook 1952), it is necessary to know the dielectric constant of a random mixture of particles in a dielectric medium. Such a dielectric theory of mixtures has been worked out mathematically by Polder and van Santen (1946), by Lewin (1947) and also by Fricke (1924) for the equivalent case of electrical conductivity. Although Fricke verified his theory by measurements of the conductivity of blood and red cell suspensions, there has been no experimental examination of the dielectric case, except on a molecular scale. We have therefore measured the dielectric constant of 5% and 10% solutions of lustrex polystyrene latex (for the loan of which we are indebted to Messrs. Monsanto Chemicals Ltd.); this is a suspension of spherical particles of polystyrene of known dielectric constant in dilute soap solution, whose dielectric constant we have also measured. This case corresponds with the type of solution for which the dielectric theory of mixtures is employed, but of course only for spherical particles.

For a small fraction x of a medium of low dielectric constant $\epsilon_{\mathrm{solute}}$ in a medium of high dielectric constant $\epsilon_{\mathrm{solvent}}$ the dielectric constant of the solution ϵ is given by all authors by the equation :

$$\epsilon_{\mathrm{solvent}} - \epsilon = \frac{\beta x}{1-x}(\epsilon - \epsilon_{\mathrm{solute}}).$$

To obtain this formula from Fricke's theory it is necessary to replace symbols for conductivity by those for dielectric constant, and also to neglect $\epsilon_{\mathrm{solute}}/\epsilon_{\mathrm{solvent}}$ in comparison with unity.

β is a parameter which Fricke has calculated for particles of spherical shape, in terms of their axial ratio. For spheres $\beta = 1{\cdot}5$, for long needles $\beta = 1{\cdot}6$ and for flat discs β rises sharply to a very high value.

Lustrex latex consists of a suspension of spherical particles of polystyrene, between 0·2 and 0·4 micron in diameter, in soap solution. The dielectric constant of the polystyrene is small compared with that of water, lying between 2·45 and 2·65 from 60 to 10^6 c/s, and it is unlikely to rise very much in the microwave region. We found the dielectric constant of the soap solution in a separate experiment, having first checked that the solution was of the same concentration as that in the latex, both by conductivity measurements using an a.c. bridge technique and by dialysis against the latex for two weeks. The particles in the latex were examined under a microscope and found to be spherical to within 5%, beyond which accuracy we could not measure. The use of more concentrated solutions was not practicable because of the aggregation of the polystyrene particles, which might affect the dielectric theory. The concentration of particles was found both by density measurements and by information provided by the manufacturer.

The dielectric measurements were made at microwave frequencies ($\lambda = 1{\cdot}26$, 3·29, 9·2 cm), where the contribution from the soap molecules is small, both because of the reduced conductivity loss and the lower relaxation frequency. But at these frequencies the water molecule itself relaxes (Collie *et al.* 1948), so that it is necessary to calculate the low frequency dielectric constant of the solutions from the Debye equations. The techniques employed, which have been described elsewhere, consist of absorption measurements in cylindrical waveguides near the cut-off region (Collie *et al.* 1948), at $\lambda = 9{\cdot}22$ cm, and a phase-balancing method in rectangular waveguides (Buchanan 1952), at $\lambda = 1{\cdot}28$ and 3·2 cm. The values of dielectric constant and loss so obtained were corrected for conductivity, which had been measured separately, and fitted to the Debye equations within $\pm 2\%$. The 'low frequency' dielectric constant of the soap solution was found to be $74{\cdot}2 \pm 0{\cdot}5$ ($\epsilon_{\text{water}} = 78{\cdot}2$), and that of the latex fell linearly with concentration to a value of $64{\cdot}7 \pm 0{\cdot}5$ at 10% polystyrene. This linear dependence on concentration is in accordance with the above theory, and it is possible from our values of $\epsilon_{\text{solvent}}$, x, ϵ and ϵ_{solute} to calculate the parameter β.

This parameter β was found to be $1{\cdot}50 \pm 0{\cdot}03$, which is in satisfactory agreement with theory. For an experimental study of particles of different shapes it would be necessary to make up a suspension of plastic spheroids in water on a macroscopic scale. There is, however, nothing in these experiments which invalidates the interpretation of a microwave dielectric measurement on aqueous solutions in terms of the mixture theories mentioned above.

Department of Physics,
University College,
London.
28th April 1953.

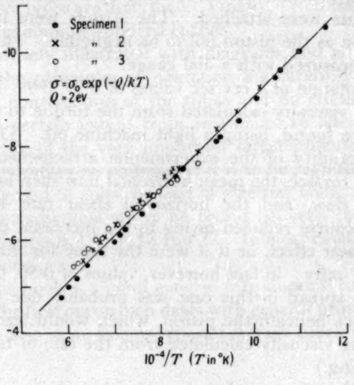
S. H. M. El-Sabeh.
J. B. Hasted.

Buchanan, T. J., 1952, *Proc. Instn. Elect. Engrs.*, Part III, **99**, 61.
Buchanan, T. J., Haggis, G. H., Hasted, J. B., and Robinson, B. G., 1952, *Proc. Roy. Soc.* A, **213**, 379.
Collie, C. H., Hasted, J. B., and Ritson, D. M., 1948, *Proc. Phys. Soc.*, **60**, 145.
Cook, H. F., 1952, *Brit. J. Appl. Phys.*, **3**, 249.
Fricke, H., 1924, *Phys. Rev.*, **24**, 575.
Haggis, G. H., Hasted, J. B., and Buchanan, T. J., 1952, *J. Chem. Phys.*, **20**, 1452.
Lewin, L., 1947, *J. Instn. Elect. Engrs.*, **94**, 65.
Polder, D., and van Santen, T. H., 1946, *Physica*, **12**, 257.

The Electrical Conductivity and Thermoelectric Power of Magnesium Oxide

Measurements of the electrical conductivity have been made on sintered specimens of magnesium oxide over the temperature range 600°c to 1500°c. The results are illustrated in the figure. The specimens were sintered at 1500°c for 14 hours, and had a density greater than 70% of the crystal density calculated from x-ray measurements. To reduce the contact resistances between electrodes and specimen platinum was evaporated on to the specimens. The results for specimens 1 and 2 were obtained by measuring the resistance between the electrodes directly, and the conductivity of specimen 3 was determined by a potential probe method. The thermoelectric power was measured by forming a pellet of MgO between two Pt : Pt–Rh thermocouples. A temperature gradient was produced by a short length of platinum wire fused on to one of the thermocouples. It was found that the thermoelectric power was positive and had a value of $1{\cdot}1 \pm 0{\cdot}1$ mv/°c over the temperature range 950°c to 1400°c.

Lempicki (1953), using single crystals of MgO, found that the conductivity was considerably reduced when the specimens were heated *in vacuo*. The values of Q in the expression $\sigma = \sigma_0 \exp(-Q/\boldsymbol{k}T)$ were between 0·15 and 0·25 ev before and in the range 2·2–2·3 ev after heat treatment. The agreement between the latter value of Q and that obtained by the author suggests that the nature of the conductivity in Lempicki's specimens after heat treatment was the same as that of the specimens used by the author. Thus the possibility that intrinsic or ionic conduction occurs in MgO is unlikely since Lempicki obtained values of the conductivity 100 times smaller than those obtained by the author. This difference can be readily explained if it is assumed that there is a stoichiometric excess of oxygen, and that the heat treatment *in vacuo* of the specimens used by Lempicki resulted in a decrease in the excess of oxygen. Confirmation that MgO is a defect semiconductor is obtained from the positive sign of the thermoelectric power.



For an impurity semiconductor two possible expressions have to be considered for the concentration of the charge carriers n (Nijboer 1939),

$$n = n_b^{1/2}\, 4{\cdot}9 \times 10^7 T^{3/4} \exp(-\Delta E/2\boldsymbol{k}T), \qquad n \gg N \quad \ldots\ldots (1)$$

$$n = (n_b/N)\, 2{\cdot}4 \times 10^{15} T^{3/2} \exp(-\Delta E/\boldsymbol{k}T), \qquad N \gg n \quad \ldots\ldots (2)$$

where $N + n_b$ is the concentration of impurity centres, and n_b is the concentration of electrons associated with them. ΔE is the activation energy of the impurity centres. The low value of the conductivity requires a very small concentration of charge carriers, thus at 600°c, $\sigma = 3 \times 10^{-11}$ ohm^{-1} cm^{-1}, and if the mobility $v > 1$ cm^2 v^{-1} sec^{-1}, then $n < 2 \times 10^8$ cm^{-3}. However, the large value of ΔE obtained using eqn. (1) indicates that only a very small fraction of the impurity centres are ionized ($T = 600$°c, $n = 2 \times 10^8$ cm^{-3}, eqn. (1) gives for $\Delta E = 4$ ev, $n_b = 10^{20}$ cm^{-3}). It would be improbable that the concentration of Schottky defects N could be less than n with such relatively large value of n_b.

It can be concluded therefore that magnesium oxide is a defect semiconductor, and that the variation of the concentration of holes should be given by equation (2).

Bedford College, London. R. Mansfield.
4th May 1953.

Lempicki, A., 1953, *Proc. Phys. Soc.* B, **66**, 281.
Nijboer, B. R. A., 1939, *Proc. Phys. Soc.*, **51**, 575.

The Variation of Transverse Viscosity with Shear Rate

A dependence of viscosity on shear rate has been detected for lubricating oils (Neale 1943, Ward, Neale and Bilton 1951), but the question of how the viscosity for a relatively small flow transverse to the main flow will be affected in such a case does not seem to have been examined. The present note gives preliminary results of work in which both viscous coefficients were measured simultaneously.

The viscometer used was a combination of the falling and rotating concentric cylinder types. A piston moved under gravity down a vertical rotating cylinder filled with oil, being restrained from rotating by a torsion arm to which strings tensioned by weights were attached. The strings were long enough for the change in inclination as the piston fell to be negligible. The vertical movement of the piston was measured with a dial gauge.

With a rate of rotation of 1 rev/sec values of 0·94, 0·99, 1·00, 1·00, 1·04, 1·08 for the ratio of the viscosity calculated from the torque to that calculated from the rate of fall were found, using a light machine oil. The spread is entirely explicable by the crudity of the experimental arrangements. When the rate was increased to 25 rev/sec, the mean horizontal flow then being about 800 times the mean vertical flow, and the horizontal shear rate $4{\cdot}4 \times 10^4$ sec^{-1}, both viscosities were of course reduced owing to the increased development of heat. If there were no shear effect, or if it were the same for the two viscosities, the ratio would remain unity: in fact however, values of 0·50, 0·59, 0·59, 0·62, 0·64 were found. (The spread in this case was probably due partly to centrifugal force in the oil space beneath the piston, which would lead to an overestimate of about 10% in the viscosity calculated from the rate of fall in the case of the lightest piston loading.)

The published data suggest that the reduction, due to shearing, in viscosity of the horizontal flow would be not more than 25%, and if so the effect of shearing on the other viscosity must be to increase it by about 25%; this seems surprising and it is planned to make more precise experiments on these lines to confirm it.

Alternative explanations of the effect are (i) differential expansion of piston and cylinder, (ii) turbulent flow, (iii) effect of temperature variation through the oil film on the two flows which are of different types—linear velocity profile and parabolic velocity profile. It is hoped to publish these considerations in more detail later, but it may be said here that none of these explanations seems tenable.

Research and Development Branch, J. W. Bremner.
Windscale Works,
Sellafield, Cumberland.
1st May 1953.

Neale, S. M., 1943, *Phil. Mag.*, **34**, 577.
Ward, A. F. H., Neale, S. M., and Bilton, N. F., 1951, *Brit. J. Appl. Phys.*, Suppl. No. 1, 12.

REVIEWS OF BOOKS

Storage Tubes and Their Basic Principles, by M. KNOLL and B. KAZAN. Pp. xiii + 143. (New York: John Wiley; London: Chapman and Hall, 1952.) 24*s*.

This book deals with the growing subject of electronic devices in which information is stored as a pattern of charge on a surface.

The book is divided, broadly speaking, into two parts. The first (43 pages) deals with the principles of operation of the charge controlled storage tubes, i.e. with the equilibrium potential distributions acquired by the insulating surfaces under various conditions of electron bombardment and secondary emission. Further the methods of writing (establishing the charge distribution) and reading (obtaining output signals from the distribution) are discussed. All special technical terms are carefully defined. In the second part (86 pages) the various kinds of storage tubes used in television, computers etc. are reviewed one by one, and the principles and performance of each are briefly described. In all some twenty-five types are described. Finally, a rich classified bibliography is provided containing about a hundred entries.

The book is profusely illustrated with diagrams and is printed in reduced typescript. While the principles of operation are carefully described, little is said of the physical construction of the tubes and no circuits are given. The volume suffers to a certain extent from references to later parts of the text, e.g. in the explanation of the very important target potential-shifting diagram (p. 6). Although this causes no ambiguity, it is to be regretted that the energies of electrons are expressed in volts, instead of electron volts.

A number of contradictions are to be found between the diagram captions and the text, e.g. in figures 4, 13, 17, 18 and 27. In most cases these can be traced to errors in the captions.

On the whole this book is a good concise introduction to the subject. This is especially true of the first part which deals with general principles. The second part will of necessity lose value as new types of storage tubes appear.

A. FOLKIERSKI.

Étude générale de l'écoulement d'un gaz a travers une tuyère quelconque et du passage par la vitesse du son (en régime permanent ou quelconque avec apport de chaleur et réaction chimique éventuels), by MAX SERRUYS. Pp. iii + 62. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air (No. 272), 1952.) 750 fr.

An Annotated Bibliography of Selected References on the Solid-State Reactions of the Uranium Oxides, by S. M. LANG. National Bureau of Standards Circular 535. Pp. iv + 95. (Washington, D.C.: U.S. Department of Commerce, 1953.) 30 cents.

Température et équilibre thermique dans une flamme de diffusion, by PIERRE BARRET. Pp. iii + 114. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air (No. 273), 1952.) 1,200 fr.

CONTENTS OF SECTION A

PAGE
Prof. P. B. Moon and Dr. A. Storruste. Resonant Nuclear Scattering of ^{198}Hg Gamma-Rays . . . 585
Dr. J. Neufeld. Ionization and Excitation Losses of Charged Particles of Intermediate Energies . . . 590
Dr. G. N. Fowler and the late Dr. G. M. D. B. Jones. On the Ionization Loss of a Fast Particle in a Dielectric Medium . . . 597
Mr. M. J. Buckingham. The Interaction of Electrons with Lattice Vibrations: Radiation by a Fast Electron . . . 601
Mr. P. Marin, Mr. G. R. Bishop and Dr. H. Halban. The Absolute Standardization of the 2·615 mev γ-Rays of ThC$''$ and the Cross Section for the Photodisintegration of the Deuteron at this Energy . . . 608
Dr. R. F. Barrow and Dr. A. D. Caunt. The $^2\Sigma^+$–$^2\Pi_1$ Band System of HBr$^+$. 617
Dr. E. W. Lee. The Influence of Domain Structure on the Magnetization Curves of Single Crystals . . . 623
Dr. E. W. T. Richards. Ionization in Liquids due to Alpha Particles . . 631
Dr. R. Wilson. A Formula for Thick Target Bremsstrahlung . . . 638
Dr. R. Wilson. Analysis of Photonuclear Reactions . . . 645

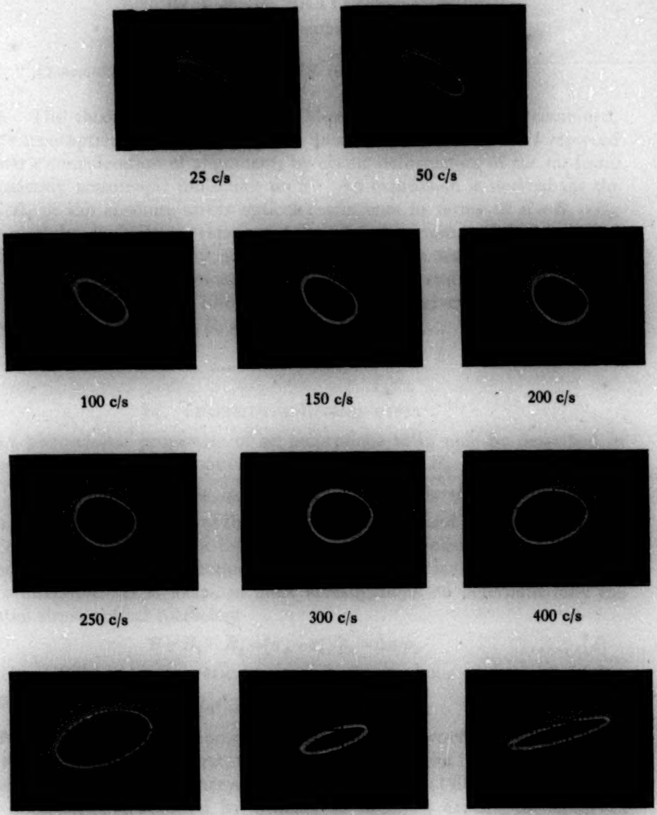
Research Notes:

Dr. H. R. Post. Many-Particle Systems: Derivation of a Shell Model . 649
Prof. C. A. Coulson. Free-Electron Wave Functions for Conjugated Molecules . . . 652
Prof. R. W. Ditchburn and Mr. G. V. Marr. The Continuous Absorption of Light in Magnesium Vapour . . . 655

Letters to the Editor:

Mr. Vachaspati. The Quantum Mechanical Equations of Motion and the Commutation Relations . . . 657
Mr. P. N. Cooper, Dr. V. S. Crocker and Dr. J. Walker. The Relative Stopping-Power of Hydrogen and of Helium for Slow α-Particles . . 658
Mr. P. N. Cooper, Dr. V. S. Crocker and Dr. J. Walker. Range–Energy Data from the ^{10}B(n, α)^{7}Li and ^{6}Li(n, t)^{4}He Reactions . . . 660
Mr. J. J. Wilkins and Mr. F. K. Goward. Alpha-Emitting Levels of ^{8}Be with Isotopic Spin $T=1$ . . . 661
Dr. P. T. Landsberg. Semiconductor Statistics . . . 662
Dr. R. H. Sloane and Mr. R. M. Hobson. The Liberation of Positive Ions by Negative Ion Bombardment of Surfaces . . . 663
Prof. S. Devons. Charge Independence and Nuclear Reactions . . . 665
Mr. K. D. Bowers. Paramagnetic Resonance in an Argentic Compound . 666
Reviews of Books . . . 667
Contents for Section B . . . 672

PROC. PHYS. SOC. VOL. 66 PT. 7—B (A. D. STUCKES)



25 c/s

50 c/s

100 c/s

150 c/s

200 c/s

250 c/s

300 c/s

400 c/s

800 c/s

1600 c/s

3000 c/s

Fig. 10. Typical set of oscillograph traces.

The Mechanism of Stress Birefringence in Amorphous Solids

By J. E. H. BRAYBON
The University, Southampton

Communicated by S. Weintroub; MS. received 30th March 1953

Abstract. The theoretical treatments of Mueller and Treloar are combined. Mueller's assumption of molecular optical anisotropy due to strain is replaced by Treloar's consideration of alignment, by elastic deformation of the medium, of permanently anisotropic molecular units. An expression is derived for the ratio B_1/B_2 of the absolute stress optical coefficients in terms of $B=B_1-B_2$, Young's modulus E, Poisson's ratio σ and the refractive index n of the unstressed material. Specimens of CR39 and Marco Resin were tested experimentally and the results of the measurements made of B_1/B_2, B and n at various wavelengths within the range 4000–7000 Å, and of E and σ are given. The theoretical and experimental values of B_1/B_2 are compared.

---

§ 1. THEORETICAL INTRODUCTION

The photoelastic properties of an amorphous solid may be described in terms of the two stress optical coefficients, B_1 and B_2, defined by

$$B_1=(n_x-n)/p_z=\delta n_x/p_z \quad \text{and} \quad B_2=(n_z-n)/p_z=\delta n_z/p_z, \qquad \ldots\ldots(1)$$

where n is the refractive index of the unstressed solid, n_x and n_z are the refractive indices of the stressed solid for light oscillating with its electric vector in the x (perpendicular to z) and z directions respectively, and p_z is the applied pressure in the z direction. It is usual, in most investigations, to determine only the differential stress optical coefficient

$$B=B_1-B_2=(n_x-n_z)/p_z=\Delta n/p_z. \qquad \ldots\ldots(2)$$

Mueller (1935), in his theoretical treatment of amorphous solids, assumed that n was given by

$$3(n^2-1)/(n^2+2)=4\pi \mathbf{P}, \qquad \ldots\ldots(3)$$

where $\mathbf{P}$ is the electric polarization, and that the following two effects were produced by the elastic deformation: first, that the Lorentz–Lorenz interaction between the dipoles altered in the manner suggested by Havelock (1908) and second, that optical anisotropy, introduced by Mueller in order to explain the observed birefringence of certain glasses, occurred. When p_z is positive, B as defined above is positive as a result of the first effect and negative as a result of the second.

The birefringence produced solely by the alignment of a molecular system, the individual units of which are permanently anisotropic, was treated theoretically by Treloar (1947) and Crawford and Kolsky (1951). Treloar, who was concerned with the long chain molecular structure of rubber (see also Kuhn and Grün 1942), considered a network of N molecular chains per cm^3, each chain consisting of a number of universally jointed links which were optically anisotropic and characterized by polarizabilities α_1 and α_2 corresponding to directions along and perpendicular to the links. Crawford and Kolsky discussed the birefringence produced by the spatial orientation under strain of a system of uniaxial rod-shaped crystallites embedded in an elastic medium. However, unless the anisotropy of the Lorentz–Lorenz interaction forces is also taken into account, the treatments lead to values of δn_x and δn_z, and hence B_1 and B_2, which must differ in sign and this is not always in accord with experimental observation.

It is possible, however, as is shown in the Appendix, to combine Mueller's and the more recent approaches to the problem. Mueller's assumption of molecular optical anisotropy due to strain is replaced by the assumption of alignment, by elastic deformation of the medium, of permanently anisotropic molecular units. Since for small strains there is little difference between the birefringence–strain curves obtained by either the Treloar or Crawford and Kolsky treatments, only the former treatment has been considered. The following relations for B_1/B_2 and B are derived, viz.

$$B_1/B_2 = \delta n_x/\delta n_z$$

$$= [M(1-2\sigma)+(K-L)(1+\sigma)]/[M(1-2\sigma)-2(K-L)(1+\sigma)] \quad \ldots\ldots(4)$$

$$B = 3(K-L)(1+\sigma)/E, \quad \ldots\ldots(5)$$

where

$$3K = 4\pi N(n^2+2)^2(\alpha_1-\alpha_2)/45n, \quad \ldots\ldots(6)$$

$$L = (n^2-1)^2/15n, \quad \ldots\ldots(7)$$

$$M = (n^2-1)(n^2+2)/6n, \quad \ldots\ldots(8)$$

and σ and E are Poisson's ratio and Young's modulus respectively. The factor K arises from consideration of the alignment of the molecular network, L from the distortion of the Lorentz–Lorenz forces and M from the change in density.

It should be noted that for rubber-like materials at small strains $\sigma \to 0{\cdot}5$ and the terms in M will disappear. B is independent of M but, because of the presence of L, the observed birefringence will tend to be less than that obtained by calculation solely on the molecular alignment theory. In fact, Treloar's value for the theoretical maximum attainable birefringence for rubber was approximately three times too large.

In order to test the validity of relations (4) and (5), B, n, E and σ were measured for two commonly used photoelastic substances, the unsaturated polyester resins CR39 and Marco Resin. The values of B_1/B_2 were then calculated by elimination of the unknown factor $K-L$ from (4) and (5) and compared with the observed values of this factor. The data are summarized in the table. Measurements were also made of the variation of B with wavelength λ because if the dispersion of B with λ is known the variation of K and thus $N(\alpha_1-\alpha_2)$ with λ may then be determined.

§ 2. EXPERIMENTAL WORK

For CR39, B was measured over the wavelength range 0·45–1·8 μ by the modified form (Filon and Harris 1931) of Filon's method and for the infra-red region a recording spectrometer (Fochs, unpublished) was used. The refractive index n was measured at four wavelengths (the intense lines, 4358, 4915, 5460, 5790 Å of a mercury vapour lamp) in the visible region by means of a CR39 prism, and the dispersion curve for the range 0·45–0·9 μ was found by the author's method (Braybon 1950) with the use of a Michelson interferometer. The ratio B_1/B_2 was obtained by the same method, and E and σ for several specimens of the material were measured by standard methods.

From the results on CR39 it appeared that a reasonable test of the theoretical relationships derived above could be obtained from measurements restricted to wavelengths in the visible region and accordingly the subsequent measurements on Marco Resin were confined to that region. For the measurement of B on this material a Babinet compensator method (Braybon and Jerrard 1952) was also used, in addition to the Filon and Harris method.

For both CR39 and Marco Resin the values of B_1/B_2 obtained by use of the Michelson interferometer had to be corrected for change in thickness of the specimen by loading. All the measurements on Marco Resin were made on specimens prepared from the same melt and reproducible results were obtained on specimens of CR39 prepared from two different melts.

Some similar measurements were made on specimens of Catalin but the results were not sufficiently reliable to permit inclusion and discussion.

§ 3. DISCUSSION OF THE RESULTS

The experimental and corresponding theoretical values are compared in the table. It will be seen that the agreement between the calculated and observed values of B_1/B_2 is good for CR39 and fair, though within the limits of the experimental error, for Marco Resin. For both materials the variation of $N(\alpha_1-\alpha_2)$ was found to be linear with $1/\lambda^2$.

Results for CR39 and Marco Resin

λ (Å)	$B\times10^{-6}$(bar^{-1})	n	$N(\alpha_1-\alpha_2)\times10^{-1}$	B_1/B_2 calc	B_1/B_2 exp
CR39 : $E=2\cdot14\times10^4$ bars, $\sigma=0\cdot37$					
5000	3·40	1·5048	2·47	1·68	1·70 ± 5%
5500	3·37	1·5018	2·45	1·68	1·70 ± 5%
6000	3·34	1·4994	2·43	1·68	1·70 ± 5%
6500	3·23	1·4976	2·42	1·68	1·70 ± 5%
7000	3·21	1·4963	2·41	1·68	1·70 ± 5%
Marco Resin : $E=1\cdot90\times10^4$ bars, $\sigma=0\cdot36$					
5000	2·25	1·5787	2·76	1·25	1·29 ± 5%
5500	2·03	1·5723	2·70	1·22	decreasing
6000	1·91	1·5672	2·64	1·21	with
6500	1·85	1·5638	2·61	1·21	wavelength to
7000	1·83	1·5603	2·58	1·21	1·28 ± 5%

Note. The stress optical coefficient in Brewsters is given by $B\times10^7$.

It is difficult to ascribe a precise physical meaning to the factor $N(\alpha_1-\alpha_2)$ in the case of cross-linked polymers such as CR39 and Marco Resin. However, since for both substances the variation of $N(\alpha_1-\alpha_2)$ with wavelength obeys a simple Cauchy relationship, it would appear likely that the natural frequencies of the anisotropic components of the molecular network lie within the ultra-violet region and not, as might at first be expected, within the infra-red region.

APPENDIX

The Change in Refractive Index of an Amorphous Solid under Unidirectional Stress

The dispersion equation

$$3(n^2-1)/(n^2+2)=4\pi\mathbf{P}, \qquad \text{......(A 1)}$$

is based on the assumption that the electric field intensity **F** acting on a macroscopic particle in an isotropic medium is

$$\mathbf{F}=\mathbf{E}+4\pi\mathbf{P}/3, \qquad \text{......(A 2)}$$

where **E** is the electric field intensity of the light wave and **P** the polarization.

The effect of a strain in the medium was considered by Mueller (1935). The strain produced a deformation of the molecular structure and hence a change in the Lorentz–Lorenz interaction between the particles. It was shown that for a medium strained in the z direction, where the strain ellipsoid has the axial ratios $1:1:1+\epsilon$, the refractive index n_x is given by

$$3(n_x{}^2-1)=4\pi N\alpha[(n_x{}^2+2)+k_x(n^2-1)], \qquad \text{......(A 3)}$$

and similarly for n_y and n_z, where $k_x=k_y=2\epsilon/5$; $k_z=-4\epsilon/5$.

In addition, changes in N and α are produced by the strain and these must be taken into account in order to obtain the full effect on n_x, n_y and n_z.

The corresponding change, δN, in N for a deformation denoted by ϵ is given by

$$\delta N=-N\epsilon. \qquad \text{.......(A 4)}$$

To calculate the change in α, Mueller's formal treatment may be replaced by Treloar's treatment, which considers the optical properties of a medium, consisting of N statistically kinked long chain molecules per cm^3. Each molecule is composed of optically anisotropic links characterized by polarizabilities α_1 and α_2 along and at right angles to their lengths. A strain in such a medium produces, solely by alignment of the molecule system, changes in the polarizability per unit volume given by

$$\left.\begin{aligned}\delta P_x=\delta P_y&=[-2(\alpha_1-\alpha_2)N/15]\epsilon,\\ \delta P_z&=[4(\alpha_1-\alpha_2)N/15]\epsilon.\end{aligned}\right\} \qquad \text{......(A 5)}$$

If eqn. (A 3) be differentiated, substitution of the values given by (A 4) and (A 5) will lead to

$$\left.\begin{aligned}&\frac{\delta n_x}{\delta\epsilon}=\frac{\delta n_y}{\delta\epsilon}=\frac{(n^2-1)^2}{15n}-\frac{(n^2-1)(n^2+2)}{6n}-(4\pi N/3)\frac{(n^2+2)^2(\alpha_1-\alpha_2)}{45n},\\ &\text{and}\\ &\frac{\delta n_z}{\delta\epsilon}=-\frac{2(n^2-1)^2}{15n}-\frac{(n^2-1)(n^2+2)}{6n}+2(4\pi N/3)\frac{(n^2+2)^2(\alpha_1-\alpha_2)}{45n}.\end{aligned}\right\} \text{..(A 6)}$$

Equations (A 6) give the Neumann strain optical constants p and q as defined by $\delta n_x=-n^2p\epsilon$ and $\delta n_z=-n^2q\epsilon$, where δn_x and δn_z are refractive index changes produced by the strain ϵ in the z direction. The relation between p and q and the constants B_1 and B_2 is $p-q=EB/n^2(1+\sigma)$; $2p+q=E(2B_1+B_2)/n^2(1-2\sigma)$.

Equations (A 6) may thus be rewritten as

$$\left.\begin{aligned} B_1 &= [M(1-2\sigma)+(K-L)(1+\sigma)]/E, \\ B_2 &= [M(1-2\sigma)-2(K-L)(1+\sigma)]/E, \\ \text{and} \quad B &= B_1 - B_2 = 3(K-L)(1+\sigma)/E, \end{aligned}\right\} \quad \ldots\ldots(\mathrm{A}\,7)$$

where $M=(n^2-1)(n^2+2)/6n,\ L=(n^2-1)^2/15n,$

and $K=(4\pi N/3)(n^2+2)^2(\alpha_1-\alpha_2)/45n.$

ACKNOWLEDGMENTS

The author wishes to acknowledge the advice and assistance of Professor A. M. Taylor, in whose department the work was performed, and to thank Mr. S. Weintroub for his help in the preparation of the manuscript.

REFERENCES

BRAYBON, J. E. H., 1950, *Proc. Phys. Soc.* B, **63**, 446.
BRAYBON, J. E. H., and JERRARD, H. G., 1952, *J. Sci. Instrum.*, **29**, 194.
CRAWFORD, S. M., and KOLSKY, H., 1951, *Proc. Phys. Soc.* B, **64**, 119.
FILON, L. N. G., and HARRIS, F. C., 1931, *Proc. Roy. Soc.* A, **130**, 410.
HAVELOCK, T. H., 1908, *Proc. Roy. Soc.* A, **80**, 31.
KUHN, W., and GRÜN, F., 1942, *Kolloid Z.*, **101**, 248.
MUELLER, H., 1935, *Physics*, **6**, 179.
TRELOAR, L. R. G., 1947, *Trans. Faraday Soc.*, **43**, 277.

Cuprous Oxide Rectifier Characteristics

By J. LEES
Department of Natural Philosophy, University of Aberdeen

Communicated by P. T. Landsberg; *MS. received* 18*th February* 1953

Abstract. Using a substitution bridge method, the current voltage and reverse capacitance voltage characteristics of a cuprous oxide rectifier have been obtained at 0°c, 30°c and 70°c. The capacitance characteristic is found to be temperature independent in the above range. The assumption of a Schottky barrier (or an allied class of barriers) is not in agreement with the experimental results. Using the capacitance characteristic and a single adjusted value of the dielectric constant of cuprous oxide, it is found possible, with only one temperature dependent parameter, to predict theoretical reverse current voltage relations in good agreement with experiment.

---

§ 1. Introduction

THE present position in the diffusion theory of metal rectifiers may be summarized as follows:

(i) *The Direct Current–Voltage Characteristic*

The shape of theoretical characteristics is not greatly influenced by the assumed impurity centre distribution. Thus a given experimental characteristic may be explained by assuming a Mott barrier, or a Schottky barrier (Landsberg 1951 a). In an explanation of this sort, however, values of various parameters of the barrier (mobilities, conductivities etc.) must be used which depend critically on the assumed impurity centre distribution; moreover, with any impurity centre distribution so far investigated these parameters are, in general, different from those known for the corresponding bulk semiconductor (Rose and Spenke 1949, Landsberg 1951 b).

(ii) *The Capacitance–Voltage Characteristic*

This depends critically on the nature of the impurity distribution. Thus a Mott barrier leads to a voltage-independent capacitance per unit area C, while a Schottky barrier (Schottky 1942), and a more general class of barriers (Landsberg 1951 b), lead to a linear relation between C^{-2} and the applied voltage. This more general class of barriers will henceforth be described as the 'class A' type of barrier. Comparisons of theory and experiment either have been confined to a limited range of applied voltage (Schottky 1942, Kobayashi and Arita 1949), or, as in the present investigation, have shown that, on the basis of known barriers, a satisfactory account of the capacitance voltage relation over a wide voltage range cannot be given (Hoffmann 1950).

Landsberg (1951 b) has shown, however, that there exists a relation, which is *independent* of impurity centre distribution in the barrier, between the d.c.

and capacitance voltage characteristics of a rectifier in the reverse direction. This relation is contained in the following equations

$$\frac{j}{j_0} = y^{3/2} e^{y} \left[1 - \exp \frac{U}{\boldsymbol{k}T} \right] \qquad \ldots\ldots(1)$$

and

$$y^2(U) - y^2(0) = 4\pi \left(\frac{\boldsymbol{e}}{\epsilon \boldsymbol{k}T} \right)^2 \int_U^0 C\,dU \qquad \ldots\ldots(2)$$

where j is the current density, ϵ is the dielectric constant, $\boldsymbol{k}T$ has its usual significance and $y^2(0)$ is a quantity which may be taken to be zero for reverse voltages U in excess of about 3 volts. Equations (1) and (2) imply that, having assigned values to ϵ and j_0, one may derive a theoretical reverse direct current voltage characteristic from the experimental reverse capacitance voltage characteristic (or vice versa).

j_0 and $y(U)$ are related to the parameters of the barrier by the additional relations

$$j_0 = \left(\frac{2}{\pi} \right)^{1/2} n(d) v \epsilon \left(\frac{\boldsymbol{k}T}{\boldsymbol{e}} \right)^2 \exp \left(- \frac{V_{\mathrm{D}}}{\boldsymbol{k}T} \right) \qquad \ldots\ldots(3)$$

and

$$y(U) = \boldsymbol{e}^2 / 2\epsilon b(U) \boldsymbol{k}T. \qquad \ldots\ldots(4)$$

Here $b(U)$ is the distance, measured into the semiconductor from the interface, of the plane at which the potential energy of the current carriers is a maximum, v is the mobility of the current carriers, $n(d)$ is the impurity centre concentration in the bulk semiconductor, and V_{D} is the 'diffusion potential'.

In this paper we describe the determination of the d.c. and reverse capacitance voltage characteristics of a single cuprous oxide rectifier at 0°c, 30°c, and 70°c. Using these results we confirm Landsberg's relation, and discuss (§4) the values of the barrier parameters which are required for agreement. For the sake of comparison the paper also contains an analysis of the experimental material on the basis of a Schottky or class A type of barrier.

§ 2. Experimental Procedure and Results

(i) *Choice of Rectifier*

Alternating and direct current measurements of metal rectifiers are complicated by creep phenomena. The creep is of a particularly simple form in cuprous oxide rectifiers, since in this type it becomes negligible within a reasonably short time of the application of a given bias voltage (about 30 minutes at 70°c, and about 30 hours at 0°c). A cuprous oxide rectifier was therefore chosen for our experiments, and at each bias voltage all the electrical measurements were made after the rectifier had reached this steady state. No creep was observed when forward bias voltages were applied to the rectifier.

Our experiments were performed on a used Westinghouse disc which had good rectifying properties and very stable characteristics. Its geometrical contact area was 2·41 cm², and the thickness of the cuprous oxide film was found by direct measurement to be about 0·014 cm. The conductivity of the bulk semiconductor was found from these dimensions and from its resistance derived from the linear part of the forward d.c. characteristic (see table 2(*c*)).

(ii) *The Rectifier Mounting*

The rectifier was mounted in a brass calorimeter placed in a water bath, the temperature of which was controlled to ± 0·1°c. The calorimeter was evacuated through a P_2O_5 water vapour trap to minimize the effects of changes in the atmosphere surrounding the rectifier. The base of the calorimeter was in good thermal contact with the copper electrode of the rectifier, and by use of a thermocouple junction, it was verified that the temperature of the copper–cuprous oxide junction did not deviate by more than 0·1°C from the mean temperature of the water bath. The counter electrode consisted of a lead washer making contact with a colloidal graphite layer painted on the cuprous oxide surface of the rectifier. Spring washers were used to keep the counter electrode in position, and to ensure that the contact pressure of the electrodes on the rectifier did not vary excessively with temperature.

(iii) *Electrical Measurements*

Bias voltages up to a maximum of 18 V in the reverse direction were applied to the rectifier; this ensured that the current flowing through the rectifier was always less than 60 mA so that excessive heating of the barrier was avoided. After creep had finished at each reverse bias voltage, the current passing through the rectifier was observed, and also the resistive and reactive parts of the 'slope impedance' as a function of frequency in the frequency range 500–9000 c/s.

Impedance measurements were made with an alternating current substitution bridge of the type used by Schottky and Deutschmann (1929), in which a small alternating voltage is superposed on the steady d.c. bias voltages. It is necessary to keep the magnitude of this alternating 'ripple' voltage as small as possible, since the non-linear behaviour of the rectifier introduces considerable harmonic distortion in the bridge. In these experiments the 'ripple' voltage never exceeded 6 mV, and usually had a value of 4 mV (r.m.s.). Rectifier noise sets a lower limit to the magnitude of the 'ripple' voltage which can be used.

The current passing with various voltages applied in the forward direction was also observed, although no impedance measurements were made in this region, since the small barrier resistance effectively shunts the capacitance and makes its measurement extremely difficult.

(iv) *Derivation of the Characteristics*

The d.c. characteristics obtained from the current measurements at 0°C, 30°C and 70°C, are shown in fig. 1.

To obtain the capacitance of the barrier at each bias voltage and temperature, the resistive component A_f of the slope impedance $A_f - iB_f$ of the rectifier was plotted against the corresponding reactive component B_f for the various frequencies f of the applied alternating 'ripple' voltage. The plot always approximated closely to a semicircle centred on the A (i.e. resistive) axis, and cutting it in two positive values. Three experimental semicircles are shown in fig. 2. The semicircular locus diagrams imply that the impedance of the rectifier may be represented by either of the two networks of fig. 3. In both of these, the capacitance may plausibly be associated with the barrier capacitance, and the resistances with the resistances of various parts of the rectifier (Henisch 1949). We have analysed our results on the basis of both networks, and find that C_1 and C_2 do not differ greatly (fig. 4).

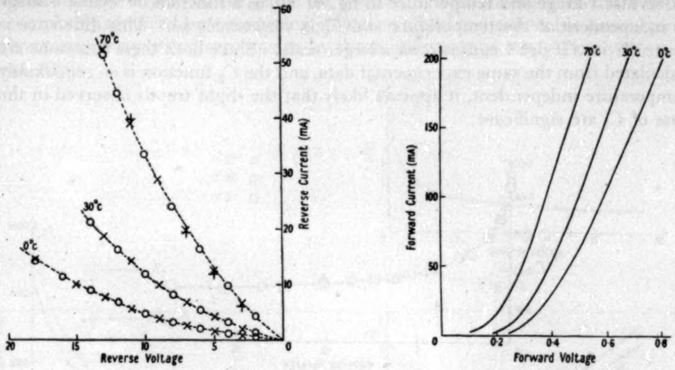
ρ is the bulk resistance of the semiconductor and is regarded as included in r_1 in the circuit of fig. 3 (*a*); in 3 (*b*) it is neglected in comparison with R_2 and r_2. C_2, R_2 and r_2 are related to C_1, R_1 and r_1 as follows:

$$C_2 = C_1(1 + r_1/R_1)^{-2}, \quad R_2 = R_1 + r_1, \quad r_2 = r(1 + r_1/R_1). \qquad \ldots\ldots(5)$$

From the semicircular diagram r_1 was obtained directly as the shorter intercept of the semicircle on the A axis, and R_2 was obtained as its diameter. C_1 was then calculated for each point A_f, B_f from the formula

$$C_1 = B_f/2\pi f(A_f - r_1)R_1. \qquad \ldots\ldots(6)$$

With large bias voltages at 70°c, the locus points tended to accumulate near the larger intercept of the semicircle on the A axis; this did not lead to any difficulties. However, with small bias voltages at 0°c the points accumulated near the smaller intercept, and, since A_f is then nearly equal to r_1, eqn. (6) became inapplicable. The capacitances were determined in these cases by equating the real and imaginary parts of the complex impedance of the circuit of fig. 3 (*a*) to those of the experimentally determined impedance at each frequency.



(*a*) (*b*)

Fig. 1 (*a*). Reverse d.c. characteristic of the rectifier: ---- experimental characteristic; × theoretical d.c. characteristic from C_1 capacitance relation; + theoretical d.c. characteristic from C_2 capacitance relation when it differs from C_1 relation; ○ theoretical d.c. characteristic on the assumption of a Schottky barrier.

Fig. 1(*b*). Experimental forward d.c. characteristic of the rectifier.

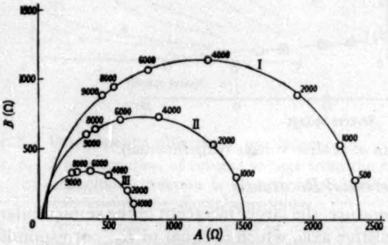


Fig. 2. Three experimental semicircular impedance locus diagrams. Figures at the experimental points indicate the frequency of measurement.

I : $T = 0$°c Bias voltage = 7·58 v
II : $T = 30$°c Bias Voltage = 2·80 v
III : $T = 70$°c Bias voltage = 0·465 v

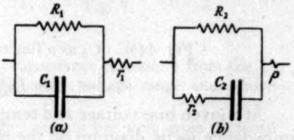


Fig. 3. Assumed equivalent circuits of the rectifier.

A small decrease in C_1 with increasing frequency was found at each bias voltage and temperature. This variation was about 10% at zero bias voltage, but was generally less than 5% from the frequency average. This average was taken to be C_1; a knowledge of the exact mechanism of the capacitance variation would probably have enabled a slightly better estimate of this capacitance to be made. C_2 was calculated by means of eqn. (5). C_1 and C_2 are shown as functions of reverse voltage and temperature in fig. 4. C_2 as a function of reverse voltage is independent of the temperature and C_1 is very nearly so. This difference is quite obvious if fig. 4 is drawn on a larger scale. Since both these functions are calculated from the same experimental data, and the C_2 function is so remarkably temperature independent, it appears likely that the slight trends observed in the case of C_1 are significant.

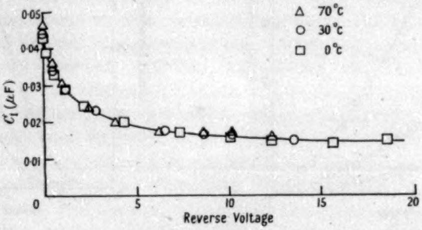


Fig. 4 (*a*). C_1 as a function of reverse voltage (experimental).

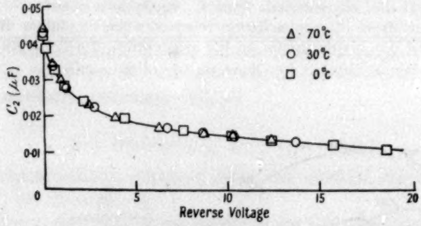


Fig. 4 (*b*). C_2 as a function of reverse voltage (experimental).

(v) *Differential Resistances*

At a given bias voltage and temperature, the larger intercept of the semicircular impedance locus diagram on the resistive axis, which is equal to R_2, corresponds to $f=0$. R_2 might therefore be expected to be equal to the slope resistance of the corresponding d.c. characteristic. Figure 5 shows R_2 and this slope resistance dU/di as functions of reverse voltage and temperature. There are considerable differences between the two functions; these are discussed in § 4 (iv).

§ 3. INTERPRETATION OF RESULTS

(i) *D.C. Relations derived from the Capacitance Relations using Equations* (1) *and* (2)

It is reasonable to assume that ϵ is temperature independent: the temperature independence of the capacitance functions lends support to this view. With this assumption we have predicted (cf. fig. 1 (*a*)) theoretical reverse d.c. relations from the capacitance characteristics using eqns. (1) and (2), which are in excellent agreement with the experimental d.c. characteristics for U greater than $3V$ (fig. 1 (*a*)). This agreement requires $\epsilon = 4{\cdot}8 \pm 0{\cdot}2$ and $\epsilon = 4{\cdot}3 \pm 0{\cdot}2$ for the C_1 and C_2 characteristics respectively. $y^2(0)$ was neglected since it is small compared with $y^2(U)$ for U greater than $3V$. As might be expected, quite large divergences (up to ~50%) were found between the experimental and predicted values of $j(U)$ in the range $0<U<3V$. These divergences could not be completely removed by introducing $y(0)$ as an additional adjustable parameter without increasing the discrepancies in the higher voltage range from 2% to about 10%.

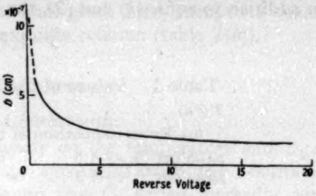


Fig. 6.

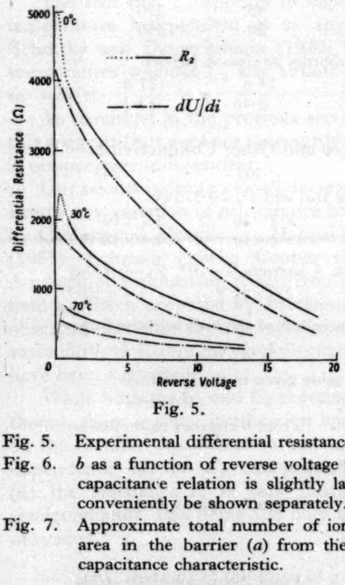


Fig. 5.

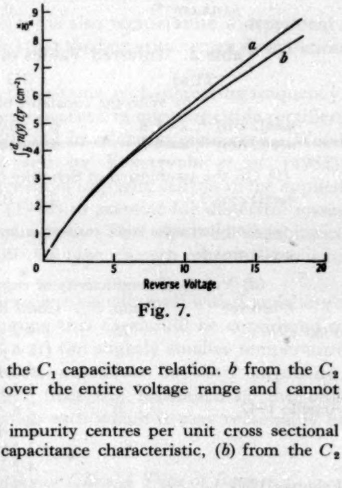


Fig. 7.

Fig. 5. Experimental differential resistances.

Fig. 6. *b* as a function of reverse voltage from the C_1 capacitance relation. *b* from the C_2 capacitance relation is slightly larger over the entire voltage range and cannot conveniently be shown separately.

Fig. 7. Approximate total number of ionized impurity centres per unit cross sectional area in the barrier (*a*) from the C_1 capacitance characteristic, (*b*) from the C_2 capacitance characteristic.

When ϵ is assumed constant in eqns. (2) and (4) the temperature independence of the capacitance characteristic implies the temperature independence of $y(U)T$ and of the function $b(U)$ which gives the distance of the potential energy maximum from the interface. In the light of eqns. (1), (2) and (4) this implies that there exists a single temperature independent function $b(U)$ from which, using one temperature independent constant ϵ, and one temperature dependent parameter j_0, the reverse capacitance and d.c. voltage characteristics of the rectifier may be satisfactorily predicted at each temperature. The function $b(U)$ obtained from the C_1 capacitance characteristic with $\epsilon = 4{\cdot}8$ is shown in fig. 6. Figure 7 gives an indication of the total number of ionized impurity centres per unit cross section of the barrier, as deduced from (Landsberg 1951 b)

$$\frac{1}{b^2} = \frac{1}{16\pi}\int_b^d n(x)\,dx. \qquad \ldots\ldots(7)$$

The broken portions of figs. 6 and 7 contain an approximation for $y(0)$ and are therefore not reliable.

(ii) *The Schottky or Class A Impurity Distributions*

The assumption of a Schottky or class A type of barrier implies that one has, in addition to eqns. (1) and (2), the relation (Landsberg 1951 a)

$$y = [(V_D - U)/V_0]^{1/4}. \qquad \ldots\ldots(8)$$

Table 1. Values of the Adjustable Scale Constants

T (°K)	273	303	343
(*a*) From comparison of capacitance and d.c. characteristics			
j_0(μA cm^{-2}), $\epsilon=4\cdot8$	11	45	220
j_0(μA cm^{-2}), $\epsilon=4\cdot3$	6·9	30	150
(*b*) On the assumption of a Schottky or class A barrier			
V_0($\times10^{-3}$ ev)	8·38	14·6	21·8
j_0(μA cm^{-2})	0·353	2·46	13·2

Table 2. Inferred Values of $n(d)ev$ and Other Conductivities

T(°K)	273	303	343
(*a*) With the constants of table 1(*a*) and $V_D=0\cdot375$ ev.			
$nev(\Omega\text{ cm})^{-1}$, $\epsilon=4\cdot8$	$6\cdot0\times10^{-3}$	$4\cdot3\times10^{-3}$	$3\cdot1\times10^{-3}$
$nev(\Omega\text{ cm})^{-1}$, $\epsilon=4\cdot3$	$4\cdot3\times10^{-3}$	$3\cdot2\times10^{-3}$	$2\cdot4\times10^{-3}$
(*b*) On the assumption of Schottky or class A barriers, $\epsilon=10$*, $V_D=0\cdot450$ev			
$nev(\Omega\text{ cm})^{-1}$	$2\cdot4\times10^{-3}$	$2\cdot0\times10^{-3}$	$1\cdot1\times10^{-3}$
(*c*) Conductivites from resistance and dimensions of the bulk semiconductor			
$nev(\Omega\text{ cm})^{-1}$	$2\cdot7\times10^{-3}$	$3\cdot2\times10^{-3}$	$3\cdot7\times10^{-3}$

(*d*) Values of conductivity of cuprous oxide given in the literature

Observer	Temp. (°c)	Cond. $(\Omega\text{ cm})^{-1}$	Remarks
Vogt 1930	75	$4\cdot2\times10^{-5}$– $9\cdot6\times10^{-3}$	From 6 cuprous oxide plates
	0	$6\cdot0\times10^{-6}$– $1\cdot5\times10^{-3}$	
Angello 1942	0	$2\cdot8\times10^{-3}$– $2\ \ \times10^{-4}$	From 7 cuprous strips
	60	$3\cdot4\times10^{-3}$– $4\cdot3\times10^{-4}$	
Feldman 1943	50	$1\cdot5\times10^{-3}$– $6\ \ \times10^{-4}$	From 2 cuprous oxide strips
Rose and Spenke 1949	30	$4\ \ \times10^{-3}$– $7\ \ \times10^{-4}$	From bulk resistances of 7 rectifiers
	30	$9\ \ \times10^{-5}$– $3\cdot5\times10^{-5}$	From barrier resistances of above rectifiers on the assumption of a Schottky barrier
Sakaki 1950	60	$1\ \ \times10^{-3}$– $7\ \ \times10^{-4}$	From 1 cuprous oxide plate

* For $\epsilon=5$ these values should be multiplied by 2.

It was found possible to choose j_0 and V_0 at each temperature to give, through eqns. (1) and (8) a theoretical reverse d.c. characteristic in good agreement with the experimental curve (fig. 1 (*a*)). j_0 and V_0 are given in table 1 (*b*). Using these values of V_0 it was found impossible, by adjusting ϵ, to obtain reasonable agreement between the values of y derived from eqn. (8), and the values of y obtained, through eqn. (2), from the capacitance characteristic using reasonable values of V_D.

We have also analysed our forward and reverse d.c. characteristics by several methods as described by Landsberg (1952), and have obtained very similar numerical results (cf. his table 1). These are not given here, except for values of the conductivity, assuming the Einstein mobility relation (table 2 (*b*)).

§ 4. CONCLUSIONS AND DISCUSSION

(i) *The Capacitance Characteristic*

The fact that C_1 appears to depend slightly on the temperature and C_2 is temperature independent is in approximate agreement with the results of Schottky and Deutschmann (1929) who found that C_1 varied markedly with temperature whereas C_2 was almost constant in the temperature range -73°C to $+20$°C.

As remarked in the previous section, $b(U)$ is also temperature independent: this suggests (but does not necessarily imply) that the impurity centre distribution is temperature independent.

Our results indicate a small decrease in capacitance with increasing frequency. Frequency variation of capacitance has been observed in cuprous oxide rectifiers by Calderwood, Cooper and Heppel (1950), and in selenium rectifiers by Wood (1933), Schmidt (1941), Cooper (1950) and by Yamaguchi *et al.* (1952). A relaxation mechanism, arising from movement of impurity centres in the applied field has been proposed by Breckenridge (1948) to account for dielectric losses observed in ionic crystals at audio frequencies. The observed capacitance variation is probably due to a mechanism of this type, though other mechanisms have been suggested.

While both the C_1 and C_2 characteristics agree equally well with Landsberg's theory, there is some justification for assuming that C_2 should be considered to be the rectifier capacitance rather than C_1: (i) the slightly smaller temperature dependence of the C_2 characteristic would otherwise be very surprising; (ii) the resistance r_1 is very much larger than the resistance of the bulk semiconductor, this being the only resistance with which it may reasonably be identified.

(ii) *Analysis of the Basis of a Schottky or Class A Type of Barrier*

The assumption of a class A type of barrier gives theoretical reverse d.c. characteristics in good agreement with experiment (fig. 1 (*a*)). Furthermore, the values of the conductivity of cuprous oxide (obtained through eqn. (3)) which are required for this agreement are not very different from those of the bulk semiconductor (tables 2 (*b*) and (*c*)) provided $V_D \sim 0{\cdot}45$ ev*. With this value of V_D it is, however, impossible to predict the capacitance characteristic.

---

* This value of V is reasonable since a study of the energy levels of cuprous oxide leads one to expect that V_D lies between 0·25 and 0·50 ev.

This characteristic can only be accounted for by assuming $V_D \sim 2$ ev which is extremely improbable and, in addition, implies impossibly large values of the conductivity.

We conclude that our experimental results cannot be analysed satisfactorily on the basis of a Schottky or class A type of barrier.

(iii) *The Values of the Parameters required to give Agreement between the Capacitance and d.c. Voltage Characteristics using Equations* (1) *and* (2)

Excellent agreement may be obtained (for $U > 3V$) between the reverse d.c. and either of the C_1 or the C_2 capacitance voltage characteristics (cf. fig. 1(*a*)). The value of the dielectric constant of cuprous oxide required for this agreement is 4·8 for the C_1 characteristic and 4·3 for the C_2 characteristic. These values are about one-half of values for the bulk material (table 3). Taking $V_D = 0{\cdot}375$ ev† values of the conductivity of cuprous oxide are implied which are of the same order as those of the bulk semiconductor but decrease as the temperature increases (table 2(*a*)).

Table 3. The Dielectric Constant of Cuprous Oxide

Observer	Temperature	Dielectric constant
Grösser and Lenertz* 1929	-93°c	12 ± 1
Højendahl 1933	21°c	10·5
Wright 1950		8

* Quoted by Schottky and Deutschmann 1929

In commenting on these results we first consider a possible error in the interpretation of our experiments. It has been suggested by Starr (1936) that the 'effective' contact area of a rectifier, as far as rectification is concerned, may be considerably smaller than the geometrical contact area. If only a fraction a of the latter is the 'effective' contact area, it can be shown that, while the agreement between theoretical and experimental curves is not altered, the values of the adjustable constants ϵ and j_0 obtained will be respectively $\sqrt{a}$ and a times their true values. Our low value of ϵ could be accounted for by taking $a = \frac{1}{4}$, the resulting change in the conductivity being insignificant in view of the uncertainty in V_D. Starr, however, inferred $a > \frac{1}{2}$ from his experimental results, and it seems unlikely that the discrepancy can be wholly accounted for thus, particularly as we might expect a satisfactory explanation to account also for the temperature dependence of the conductivities.

Secondly we must consider the assumptions made in deriving the equations of §1, and the ways in which the rectifier may violate these assumptions. (*a*) The diffusion equation and the Einstein mobility relation are assumed to be applicable. (*b*) It is assumed that the potential function in the barrier has a single sharp maximum and that the Laplace approximation (Landsberg 1951 b) is therefore applicable. This is not necessarily the case, since the discontinuous nature of the space charge distribution could imply a relatively broad maximum or even a series of comparable maxima in the potential function. (*c*) The possibility of wave mechanical penetration of the top of the barrier has not been considered. In the case of crystal rectifiers, Courant (see Torrey and Whitmer 1948) has made an estimate of this effect, and has shown that it brings about an increase in the current (at a given bias voltage) comparable with that due to the image force. An approximate calculation confirms this result, and suggests that the effect will be less important as the temperature increases. We should therefore expect that, due to 'tunnel' effect, the current flowing at a given bias voltage will be larger than that given by the theory. j_0, and our deduced conductivities, will therefore be larger than the theoretical values, but will tend towards them as the temperature increases.

---

† See footnote on preceding page.

A combination of the above effects may well account for the numerical values obtained, but their detailed explanation requires further investigation. In an additional experiment on another cuprous oxide rectifier of similar origin at 25°C, the relation embodied in eqns. (1) and (2) was again demonstrated and the values of the parameters required for agreement were found to be very similar to those obtained above.

(iv) *Differential Resistances*

To explain the difference between R_2 and dU/di (§ 2(v)) we consider fig. 8 which shows the current voltage relations of a cuprous oxide rectifier. Curve XOY represents the d.c. characteristic of the rectifier when all creep has been eliminated. If the rectifier is in the state represented by the point O and the applied bias voltage V is suddenly changed by $\pm\delta V$, then the current does not immediately assume the values given by points A or B. We have verified experimentally that this current change does, in fact, take place in two fairly distinct parts: at first there is a rapid ($\sim 10^{-5}$ second) change, represented by OC or OD—this corresponds to a change in current with the rectifier still effectively in the equilibrium state corresponding to the voltage V; it is followed by slow (of the order of several minutes) current creep, represented by CA (positive creep) or DB (negative creep), as the rectifier adjusts itself to the equilibrium conditions characteristic of the bias voltage $V+\delta V$ or $V-\delta V$ respectively. Thus R_2, which was measured with an alternating voltage of frequencies between 500 and 9000 c/s, is the differential resistance corresponding to the curve CD, whereas the d.c. 'slope' resistance corresponds to the curve AB. R_2 is therefore greater than the corresponding d.c. 'slope' resistance, and experiments indicate that the ratio IC/IA has approximately the correct value to produce the observed difference.

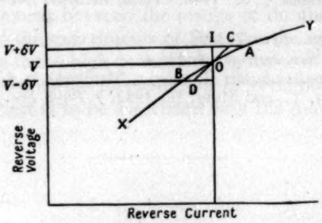


Fig. 8.

ACKNOWLEDGMENTS

I am indebted to the Department of Scientific and Industrial Research for a maintenance allowance and to Professor R. V. Jones for placing research facilities at my disposal. My sincere thanks are due to Dr. P. T. Landsberg for his guidance during this work and to Dr. C. W. McCombie for many helpful suggestions.

REFERENCES

ANGELLO, S. J., 1942, *Phys. Rev.*, **62,** 371.
BRECKENRIDGE, R. G., 1948, *J. Chem. Phys.*, **16,** 959.
CALDERWOOD, J. H., COOPER, R., and HEPPEL, H. K., 1950, *Research*, **3,** 350.
COOPER, R., 1950, *Proc. Phys. Soc.* B, **63,** 176.
FELDMAN, W., 1943, *Phys. Rev.*, **64,** 113.
HØJENDAHL, K., 1933, *Z. Phys. Chem.*, **320,** 54.
HOFFMANN, A., 1950, *Z. angew. Phys.*, **2,** 353.
HENISCH, H. K., 1949, *Metal Rectifiers* (Oxford : Clarendon Press), p. 33–35.
KOBAYASHI, A., and ARITA, K., 1949, *J. Phys. Soc. Japan*, **4,** 58.
LANDSBERG, P. T., 1951 a, *Proc. Roy. Soc.* A, **206,** 463; 1951 b, *Ibid.*, **206,** 477; 1952, *Ibid.*, **213,** 226.
ROSE, F., and SPENKE, E., 1949, *Z. Phys.*, **126,** 632.
SAKAKI, W., 1950, *J. Phys. Soc. Japan*, **5,** 455.
SCHMIDT, A., 1941, *Z. Phys.*, **117,** 754.
SCHOTTKY, W., 1942, *Z. Phys.*, **118,** 539.
SCHOTTKY, W., and DEUTSCHMANN, W., 1929, *Phys. Z.*, **30,** 845.
STARR, J., 1936, *Physics*, **7,** 16.
TORREY, H. C., and WHITMER, C. A., 1948, *Crystal Rectifiers* (New York : McGraw-Hill), p. 87.
VOGT, W., 1930, *Ann. Phys., Lpz.*, **7,** 183.
WOOD, L. A., 1933, *Rev. Sci. Instrum.*, **4,** 434.
WRIGHT, D. A., 1950, *Semiconductors* (London : Methuen), p. 42.
YAMAGUCHI, J., MIYAUCHI, T., and MORI, H., 1952, *J. Phys. Soc. Japan*, **7,** 171.

The Barkhausen Effect in Single Crystals

By R. S. TEBBLE and V. L. NEWHOUSE*
Department of Physics, University of Leeds

Communicated by E. C. Stoner ; *MS. received 24th April* 1953

Abstract. Measurements have been made on the Barkhausen effect in single crystals of silicon–iron and nickel and it is shown that the results are markedly dependent on the demagnetizing factor of the specimen. The effect of a high demagnetizing factor is to reduce the magnitude of the Barkhausen effect and to produce a corresponding increase in the contribution of the reversible magnetization processes. An explanation of this effect is given. It is pointed out that the good agreement between the results of the theoretical investigation of Néel and others and the experiments of Bates on the magnetization of single crystals is probably due to the high demagnetizing factor of the specimens, which results in a close approximation to conditions of reversibility. It is suggested that the Barkhausen effect is to be associated with the movement of 180° rather than 90° boundaries.

§ 1. INTRODUCTION

DURING the decade following the discovery of the magnetic Barkhausen effect it was believed that the volumes of magnetic material involved in a sudden change of magnetization could be identified with the domain volumes of the specimen. That this simple view is untenable has been shown only by comparatively recent work. On the one hand a somewhat more precise knowledge of domain sizes has been obtained by the study of powder patterns on single crystals (Elmore 1938, Williams, Bozorth and Shockley 1949, Bates and Neale 1950, and others); on the other, definite information has been obtained about the number and size distribution of Barkhausen discontinuities in polycrystalline specimens in the work of Bush and Tebble (1948) and Tebble, Skidmore and Corner (1950). These papers will be referred to as I and II respectively. It has become apparent that the greater part of the ordinary Barkhausen discontinuities must be associated with much smaller volumes than the primary domains and are probably produced by the irregularities in the movement of the main domain boundaries.

In the work of Tebble and his collaborators, the study of the Barkhausen effect has been linked with measurements on the reversible susceptibility so as to obtain an estimate of the relative importance of reversible and irreversible processes over the hysteresis range. It is obviously desirable that work of this nature should not be confined to polycrystalline specimens but that the investigations should be extended to single crystals in view of the considerable advances in our knowledge of their domain structure.

---

* Now with Messrs. Ferranti Ltd., Manchester.

§ 2. EXPERIMENTAL

2.1. *Pulse Counting Measurements*

The method used for the size measurement and counting of the magnetic discontinuities was essentially that described in II and therefore need only be described in brief. The specimen under investigation is contained inside the 'search-coil' which is itself situated inside the magnetizing coil, and the whole arrangement is screened by a mumetal box. During a slow change of magnetizing field the Barkhausen discontinuities produced in the specimen induce voltage pulses in the search coil; these voltage pulses are amplified, passed through a voltage discriminator and finally recorded by a suitable counting unit.

The rate of change of the magnetic field in the present work was of the order of 1 oersted per minute and it was found convenient to control the current in the magnetizing coil by means of an electrolytic cell consisting of 2% copper sulphate with copper electrodes. The resistance of the cell was changed by allowing the electrolyte to drip slowly from the container; in this way runs lasting up to 60 hours have been taken.

For counting at different pulse levels the discriminator setting was held constant, and the output attenuator of the amplifier was varied. For a particular attenuator setting readings of the counter were taken at field intervals of 2 oersteds as the specimen was taken through a portion of the magnetization curve. The maximum rate of entry of pulses into the scaling unit (resolving time 10^{-4} sec) was not allowed to exceed 70 per second.

2.2. *Determination of the Total and of the Reversible Susceptibility*

The magnetic susceptibility, referred to here as the 'total' susceptibility, was determined by graphical differentiation of the hysteresis loop which was determined by the standard ballistic method. Owing to the small size of the specimens investigated, their large demagnetizing factors and consequent small effective permeability, the correction which had to be made to the experimental readings due to the area of the induction measuring coil was unusually large. Special precautions had therefore to be taken to ensure that the induction coil had exactly the same orientation with reference to the magnetizing field both when it was empty for purposes of calibration and when it contained the specimen whose hysteresis loop was being measured. Similar precautions had to be taken to ensure that the orientation of the specimen with regard to the magnetic field was precisely the same for the hysteresis loop measurements and for the reversible susceptibility measurements for which the same coil system was used. For the hysteresis loop measurements a heavy copper tube was inserted round the specimens and inside the magnetizing coil. This was to ensure that the changes of field during the course of the measurements took place slowly, in accordance with the conditions for specimens of high demagnetizing factor given by Snoek (1939).

The reversible susceptibility of the specimens was measured by means of an improved version of the low frequency bridge method used by Tebble and Corner (1950). The definition of reversible susceptibility used in their work, which will be adopted here, is $\kappa_{\text{rev}} = \text{Lim}_{\Delta H \to 0}(\Delta I/\Delta H)$, where I is the magnetization, and ΔH is an alternating field superimposed on a steady polarizing field. The effective susceptibilities uncorrected for demagnetizing factor were so low that eddy current corrections were negligible.

The measurements obtained in the present work were actually of the effective susceptibility given by

$$\kappa = \Delta I/\Delta H_{\text{applied}} = \Delta I/\Delta(H_{\text{eff}} + NI), \qquad \ldots\ldots(1a)$$

and the measured total susceptibility κ_{tot} is given by

$$\frac{1}{\kappa_{\text{tot}}} = \frac{1}{\kappa^0_{\text{tot}}} + N, \qquad \ldots\ldots(1b)$$

where N is the demagnetizing factor and κ^0_{tot} the total susceptibility with zero demagnetizing factor.

2.3. *Specimens*

The choice of specimens in work of this nature requires careful consideration if the results which are obtained are to be of any significance. Crystals of silicon–iron and of nickel were available but it was realized that, because of the relatively small length to diameter ratio of the specimens which could be produced, their properties would be strongly dependent on their shape. It was obviously desirable that the effect of alteration in shape and in crystal structure should be examined. Silicon–iron crystals with a different shape but similar magnetic properties were produced and a polycrystalline specimen of similar dimensional ratio prepared.

The silicon–iron crystal was taken from a sheet of hot rolled silicon–iron alloy known as Stalloy II, manufactured by G.K.N. Ltd. In the form in which it is manufactured this material sometimes forms large crystals which fracture easily along cleavage planes. This specimen was shaped by producing suitable fracture in a large crystal. The silicon–iron II crystal was prepared from the silicon–iron I specimen by etching the sides of the crystal so as to reduce the cross section while the length was left relatively unaltered. Whereas the cross section of the silicon–iron I crystal was approximately square the second specimen was roughly cylindrical in shape.

The polycrystalline specimen, which was taken from a length of wire similar to the large grained iron specimen described in II, so as not to introduce unnecessary strains by cutting, was etched from the wire and was approximately ellipsoidal in form.

The nickel crystal which was produced from a larger crystal by etching, was in the form of a long plate with reasonably straight sides but a rather irregular surface.

The main properties of the specimens are summarized in table 1; the value of H_k given in the table is the field value corresponding to the 'knee' of the hysteresis curve, below which the magnetization curve is approximately linear.

§ 3. RESULTS

The discriminator transmits pulses greater than a given size and in papers I and II the results have been given as number–level (N, M) curves, showing N the number of Barkhausen discontinuities per cm^3 above a given size M, plotted against the corresponding discriminator or attenuator setting (in terms of M).

The contribution of the measured discontinuities over the measured range of M is then given by

$$I_B = \int M dN, \qquad \ldots\ldots(2)$$

with appropriate limits (cf. II, page 755; the symbol Δ has been dropped in the present paper).

In the present work the results are given in terms of susceptibility, and the 'Barkhausen susceptibility' is

$$\kappa_B = \frac{dI_B}{dH} = \frac{d}{dH}\int MdN = \int \frac{MdN}{dH} = \int MN', \qquad \ldots\ldots(3)$$

where N' is the number of Barkhausen discontinuities per unit change in field. Typical (N', M) curves for silicon–iron I and II at remanence are given in fig. 1: those for the other specimen are of a similar form. Over the range of measurement the (N', M) curves can be fitted to curves of the type $A \exp(-\kappa M^{1/n})$ and the values of the parameters A, K and n are given in table 2.

Table 1. Physical Properties of Specimens

Specimen	Composition	Length (cm)	Cross sectional area (mm²)	Direction of Long Axis	Angle of face (°)*
Nickel crystal		0·620	$0{\cdot}10_2$	(100)	(38, 52, 90)
Silicon–iron crystal I	Si 4·32 C 0·037 S 0·010 P 0·008 Mn 0·14 Ni 0·06 Cr 0·025 Al 0·022	0·688	$0{\cdot}12_0$	(201)	(26, 63·5, 63·5)
Silicon–iron crystal II		0·657	$0{\cdot}045_2$		
Polycrystalline iron	Mn 0·05, Fe 99·92	0·869	max. dia. 0·43 mm.		

* Angle between normal to broadest face and nearest (100) axes.

Table 2. Summary of Results of Magnetic Measurements

Specimen	H_c	H_k	κ_{rev}*	κ_{tot}	$1/N$	(N', M) curve parameters		
						A	K	n
Nickel crystal	5	60	5·5	9·6	9·3	$4{\cdot}5\times10^8$	17000	2·0
Silicon–iron I	1·4	130	9·7	10·5	10·0	$2{\cdot}5\times10^6$	310	3·1
Silicon–iron II	1·4	80	14·6	18·1	19	$3{\cdot}0\times10^6$	2100	2·2
Polycrystalline iron	1·5	100	7·7	12·0	11·7			

* at remanence.

An indication of the reliability of the total susceptibility κ_{tot} measurements is obtained from the fact that, for specimens with a high demagnetizing factor N, $\kappa_{tot} \simeq 1/N$. The experimental values of the total susceptibility at remanence agreed, to within 5%, with the reciprocals of the demagnetizing factors, calculated from the dimensions of the specimens.

Owing to the small size of the specimens, the reversible susceptibility measurements had to be carried out using rather large values of the alternating field. Measurements were carried out for each crystal, on the variation of the experimental susceptibility value with the amplitude of the alternating field used in its determination. Extrapolation of the curves obtained shows that in the case of the nickel crystal the experimental values of the reversible susceptibility are unlikely to be too high by more than approximately 2·5% near remanence.

In the case of the silicon–iron crystals however, the values of the reversible susceptibility, at least near remanence, may be as much as 20% too large, although the error may be much less at high fields. This uncertainty does not invalidate the conclusions which will be drawn from these results.

Curves showing the variation with field (increasing from negative to positive values) of κ_{tot}, κ_{B} and κ_{rev} for the single crystals are shown in figs. 2, 3 and 4. A summary of the results of the magnetic measurements is given in table 2. None of the results are corrected for demagnetizing factor.

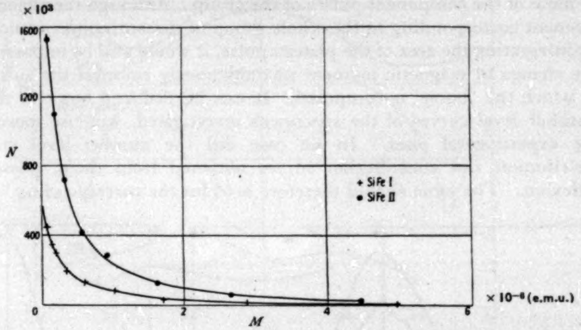


Fig. 1. (N', M) curves for silicon–iron crystals I and II at remanence, showing the number of discontinuities per cm³ per unit change in field with a magnetic moment greater than M.

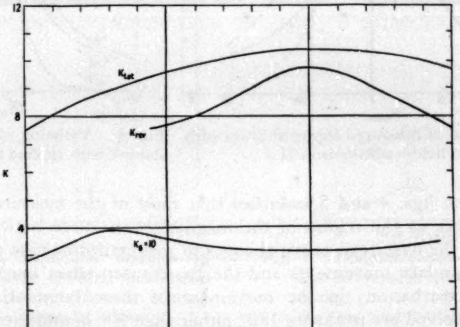


Fig. 2. Variation of measured susceptibilities with applied field—silicon–iron I.

§ 4. DISCUSSION

Examination of figs. 2, 3 and 4 shows that in no case does the integrated Barkhausen contribution (due to measured discontinuities larger than 10^{-7} e.m.u.) account for more than 30% of the difference, i.e. $\kappa_{\text{tot}}-\kappa_{\text{rev}}=\kappa_{\text{irrev}}>\kappa_{\text{B}}$ between the values of the total and the reversible susceptibility. It seems fairly certain that extrapolation to $M=0$ can do no more than halve this discrepancy. The reason for this is almost certainly due to the fact, previously mentioned in paper II, that the Barkhausen pulses occur in groups, of which only one tends to be detected by the apparatus used. It would seem to be fundamentally impossible both to count and to measure the magnetic discontinuities produced in metallic specimens since a large proportion of them occur in bursts. Because of the eddy current effect, the voltage pulse induced in a search coil due to such a burst will have the shape of a plateau with several peaks, each peak corresponding to one or more of the component pulses of the group. Although the change of magnetic moment corresponding to the whole group of discontinuities could be evaluated by integrating the area of the plateau pulse, it would still be impossible to allocate the change of magnetic moment unambiguously amongst the subsidiary pulses of which the 'burst' is composed. It can be deduced however that the 'true' number–level curves of the specimens investigated, will rise more steeply than the experimental ones. In no case did the number–level curves, or the distribution and contribution curves obtained from them, possess points of inflexion. The same should therefore hold for the corresponding 'ideal' curves.

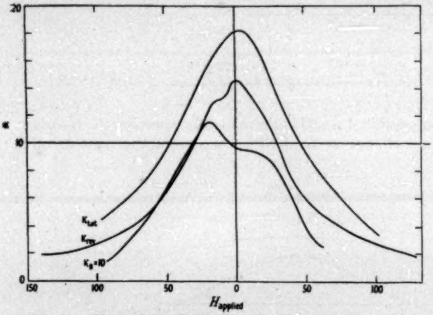


Fig. 3. Variation of measured susceptibilities with applied field—silicon–iron II.

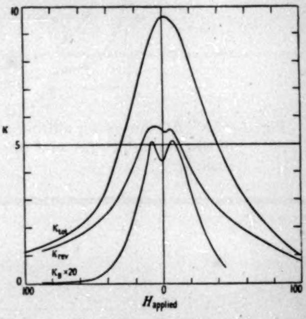


Fig. 4. Variation of measured susceptibilities with applied field—nickel crystal.

Reference to figs. 4 and 5 indicates that most of the measured Barkhausen effect takes place in the region of the magnetization curve between $H = \pm H_k$. In this part of the hysteresis curve changes in magnetization take place primarily by domain boundary movements and the Barkhausen effect must be associated with small perturbations in the movement of these boundaries. The wall movements involved are probably 180° rather than 90° boundaries as the former can 'wrinkle' without the production of 'free charge' whereas the latter cannot. This is illustrated in fig. 6. Although the presence of strains and inclusions will modify this picture considerably there will be no essential change in the concept that 180° boundaries are energetically more favourable for the production of irregularities than are 90° boundaries.

Perhaps the most striking feature of the results is the fact that such a large portion of the total susceptibility is accounted for by the reversible component. This is the case for the polycrystalline specimen as well as the single crystals. The fact that this phenomenon is probably due to the high demagnetizing factors of the specimens investigated, is illustrated in fig. 5 where the fractions $\kappa_{\mathrm{rev}}/\kappa_{\mathrm{tot}}$ and $\kappa_{\mathrm{B}}/\kappa_{\mathrm{tot}}$ for the two silicon–iron crystals and $\kappa_{\mathrm{rev}}/\kappa_{\mathrm{tot}}$ for the polycrystalline iron specimen are plotted against the reduced field H/H_{k}. For specimens of polycrystalline material having a very low demagnetizing factor $\kappa_{\mathrm{rev}}/\kappa_{\mathrm{tot}}$ is of the order of 5% and hence $\kappa_{\mathrm{irrev}}/\kappa_{\mathrm{tot}} \simeq 95\%$ (see paper II). Not only is there a reduction in the Barkhausen contribution but it is evident from an examination of fig. 1 and the parameters given in table 2 that the size of the Barkhausen discontinuities is also reduced by an increase in demagnetizing factor.

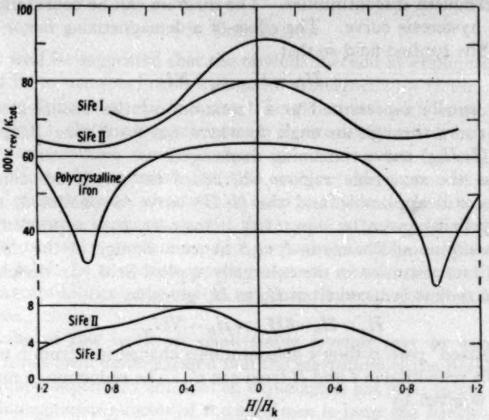


Fig. 5. Effect of shape on the reduced reversible and irreversible susceptibilities (note change of scale in bottom curves, which show $100\kappa_{\mathrm{B}}/\kappa_{\mathrm{tot}}$).

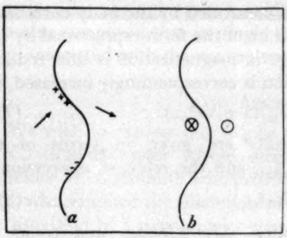


Fig. 6. (*a*) A ' forbidden ' domain wall configuration showing production of ' free charge ' (direction of magnetization shown by arrows).
(*b*) An ' allowed ' domain wall configuration (direction of magnetization normal to plane of diagram).

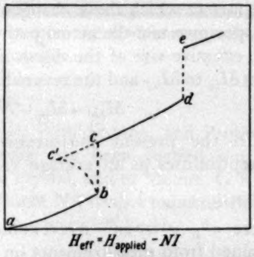


Fig. 7. Diagram representing reversible and irreversible changes in magnetization. ab, cd represent reversible changes and bc, de represent discontinuous irreversible changes.

It would seem safe to conclude from these results that the reversible proportion of the total susceptibility rises as the demagnetizing factor increases and that both the irreversible contribution and the magnitude of the Barkhausen discontinuities decrease correspondingly. This can be explained by reference to fig. 7. In a specimen with zero demagnetizing factor the sequence of reversible and irreversible changes in magnetization may be represented by a series of steps ab, bc, etc. in which ab, cd . . . represent reversible changes in magnetization and bc, de . . . represent the discontinuous changes in magnetization which produce Barkhausen discontinuities. The diagram can be considered as a much 'magnified' hysteresis curve. The effect of a demagnetizing factor is to reduce the effect of the applied field so that

$$H_{\mathrm{eff}} = H_{\mathrm{applied}} - NI, \qquad \ldots\ldots(4)$$

and this is usually represented as a 'rotation' of the uncorrected hysteresis (I, H_{applied}) curve through an angle θ, where $\tan\theta = N$ (see Bates 1951), the 'corrected' (I, H_{eff}) curve remaining unchanged.

So far as the reversible regions ab, cd of the curve are concerned this correction is still applicable and the (I, H) curve is unaltered, the effective susceptibility being given by eqn. (1 *a*). However, over a portion such as bc there is a discontinuous change in I, and hence a change in the demagnetizing field NI, with no alteration in the externally applied field. In the change bc the effective field is thus reduced from H_{b} to $H_{\mathrm{c}'}$ given by

$$H_{\mathrm{c}'} = H_{\mathrm{b}} - \delta H_{\mathrm{c}'\mathrm{c}} = H_{\mathrm{b}} - N\delta I_{\mathrm{bc}'}, \qquad \ldots\ldots(5)$$

and the 'idealized' path is thus a discontinuous change in I from b to c followed by a reversible reduction in I and H from c to c′. The reduction in magnetization from c to c′ is given by

$$\delta I_{\mathrm{c}'\mathrm{c}} = \kappa^{0}_{\mathrm{rev}}\delta H_{\mathrm{c}'\mathrm{c}} = \kappa^{0}_{\mathrm{rev}} N\delta I_{\mathrm{bc}'}, \qquad \ldots\ldots(6)$$

where $\kappa^{0}_{\mathrm{rev}}$ is the reversible susceptibility corrected for demagnetizing factor. The rate at which these changes take place is controlled by the eddy currents in the specimen and the actual path followed will be of the form represented by bc′. The effective size of the discontinuous change in magnetization is thus reduced from δI_{bc} to $\delta I_{\mathrm{bc}'}$ and the reversible contribution is correspondingly increased, i.e.

$$\delta I_{\mathrm{bc}'} = \delta I_{\mathrm{bc}} - N\kappa^{0}_{\mathrm{rev}}\delta I_{\mathrm{bc}'} = \delta I_{\mathrm{bc}}/(1 + N\kappa^{0}_{\mathrm{rev}}). \qquad \ldots\ldots(7)$$

In the present measurements the results are given in terms of the susceptibilities as an average value $\delta I/\delta H_{\mathrm{applied}}$, and the relevant expression is

$$\kappa_{\mathrm{irrev}} = \kappa^{N}_{\mathrm{irrev}}(1 + N\kappa^{0}_{\mathrm{rev}}), \qquad \ldots\ldots(8)$$

where $\kappa^{N}_{\mathrm{irrev}}$ is an 'irreversible susceptibility' $\Delta I_{\mathrm{bc}}/\Delta H_{\mathrm{applied}}$ and could be obtained from measurements on a specimen with zero demagnetizing factor and applying eqn. (1 *a*).

The factor $N\kappa^{0}_{\mathrm{rev}}$ is thus a measure of the reduction in the Barkhausen effect produced by the demagnetizing factor. Because of the high demagnetizing factor of the four specimens used it is practically impossible to make a reasonably accurate estimate of $\kappa^{0}_{\mathrm{rev}}$ but it is probable that $\kappa^{0}_{\mathrm{rev}}$ is of the order 50 to 100 at remanence giving, with $N = \frac{1}{10}$, a value of $1 + N\kappa^{0}_{\mathrm{rev}}$ between 6 and 11, which is of the order of the reduction in the magnitude of the Barkhausen effect.

The present results would seem to explain the fact that the treatments of the magnetization curve of single crystals which have been presented by Néel (1944) and Lawton and Stewart (1948), assuming conditions of complete reversibility, are in such good agreement with experiment (Bates and Neale 1950). This is presumably due to the fact that the crystals investigated experimentally have all had rather large demagnetizing factors; although the experimental conditions were such as to reduce the demagnetizing effect they would correspond more closely to the theoretical conditions of reversibility than would have been generally expected. It is hoped that a detailed treatment of the subject will be presented in a separate paper.

§ 5. CONCLUSION

It might well be suggested that the obvious method of eliminating the shape effect would be to use specimens with small demagnetizing factor. In addition to the obvious difficulties of producing a single crystal with length to diameter ratio of say 100 : 1, the consequent increase in size would result in an increase in the already considerable time involved in carrying out an analysis of the Barkhausen effect. It is possible that this problem may be solved by the careful choice of a 'picture frame' crystal, but even in this case there would still be a considerable increase in dimensions if the crystal is to be sufficiently robust to allow for the necessary machining and preparation. The simplest solution to the problem would be to confine investigations to measurements of the total and reversible susceptibilities only and to estimate the reversible contribution from the difference.

The results of this work on small single crystals may be summarized by stating that it has been demonstrated that the magnitude of the Barkhausen effect is strongly shape dependent and that an explanation has been given. It is shown that if the demagnetizing factor of the specimen is large the Barkhausen effect is small and the magnetization process is mainly reversible. This effect would explain the good agreement between some of the theoretical and experimental investigations on single crystals, which assume conditions of complete reversibility. It is also suggested that in single crystals such as iron the Barkhausen effect is to be associated with the movement of 180° boundaries.

ACKNOWLEDGMENTS

We wish to express our thanks to Professor R. Whiddington, and Professor E. C. Stoner, under whose direction this work has been carried out, to Professor W. Sucksmith who provided the nickel crystal, and to Messrs. G.K.N. Ltd., who supplied the silicon–iron. One of us (V. L. N.) wishes to acknowledge the provision of a maintenance grant from the University of Leeds.

REFERENCES

BATES, L. F., 1951, *Modern Magnetism* (Cambridge : University Press), p. 63.
BATES, L. F., and NEALE, F. E., 1950, *Proc. Phys. Soc.* A, **63,** 374.
BUSH, H. D., and TEBBLE, R. S., 1948, *Proc. Phys. Soc.*, **60,** 370.
ELMORE, W. C., 1938, *Phys. Rev.*, **53,** 757.
LAWTON, H., and STEWART, K. H., 1948, *Proc. Roy. Soc.* A, **193,** 72.
NÉEL, L., 1944, *J. Phys. Radium*, **5,** 265.
SNOEK, J. L., 1939, *Physica*, **6,** 161.
TEBBLE, R. S., and CORNER, W. D., 1950, *Proc. Phys. Soc.* B, **63,** 1005.
TEBBLE, R. S., SKIDMORE, I. C., and CORNER, W. D., 1950, *Proc. Phys. Soc.* A, **63,** 739.
WILLIAMS, J. H., BOZORTH, R. M., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75,** 155.

Oscillations in Space Charge Detectors and the Ultra-Ionization Potentials of Mercury

By O. M. WHITE

Department of Physics, Queen's University, Belfast

Communicated by R. H. Sloane; *MS. received* 3*rd March* 1953

Abstract. Experiments have been carried out on the probability of ionization of mercury by electron impact, using an electron gun and space charge detector. Some irregularities in the curve obtained for the variation of probability with electron energy have been found to be associated with instability and oscillations in the detector. These appear to arise from a periodic rearrangement of space charge. The results show that under certain conditions the response of a space charge detector can be non-linear and that it can give spurious effects which might be misinterpreted as marking the onset of new ionization processes.

---

§ 1. INTRODUCTION

SINCE the work of Lawrence (1926) a number of investigations have been carried out on the ionization of mercury vapour by electron impact, which indicate that there are irregularities in the ionization probability curve above 10·38 v. Some experiments show the irregularities as changes in slope, others as maxima and minima, and there is marked disagreement in the number reported by different observers. The true probability curve seems to have been determined by Nottingham (1939). His results indicate that the chance of ionization does not increase regularly with electron energy above the first ionization potential of 10·38 v. Instead, it passes through a maximum at 10·8 v, a minimum at 11·05 v, a second less pronounced maximum at 11·45 v, and thereafter, although probably continually increasing up to at least 12·5 v, does so irregularly. Nottingham used a magnetic analyser to reduce the velocity spread of the electron beam, and collected the positive ions by a small wire made slightly negative to the space. He used every available technique to obtain pure mercury vapour and to reduce spurious effects. His results are supported by later work by Stevenson and Hipple (1942). Accepting Nottingham's results as most probably correct, an explanation must be sought for the disagreements in the results of the earlier experiments, many of which were done with great care. McFadden's experiments (1946) were a first attempt in this direction. He obtained results apparently similar to Nottingham's, using a simple tube with a pure tungsten filament and detecting the positive ions by a space charge detector used by the Hertz method. A sharp minimum in the curve showing the variation of ionization cross section with voltage was recorded at 10·85 v. The tube was not however designed for exact measurement of fine structure and it was surprising that it showed such detail. McFadden has discussed the possibility of the irregularities arising from such effects as periodicities in the structure of the filaments and grid, or variations in thermionic emission constants of the material of the filament.

The work described in this paper is an extension of McFadden's. It has been shown that the irregularities which appeared on his ionization probability curve almost certainly arose from an instability in the space charge detector. This instability can be associated with low frequency oscillations and is probably due to a discontinuous rearrangement of space charge. It is possible that this effect may have caused some of the disagreement in the results of the earlier experiments on ionization of mercury vapour. It does not directly affect the reliability of Nottingham's results, as his experimental method was different.

§ 2. EXPERIMENTAL METHOD AND RESULTS

The arrangement of the electrodes in the experimental tube is shown in fig. 1. A pure tungsten filament F, 0·1 mm in diameter and 3·0 mm in length, was set at a distance of 2·0 mm from, and at right angles to, a slit S of width 0·5 mm to which electrons from F could be accelerated. The slit was covered with fine mesh gauze, the wires of which were set diagonally to reduce possible effects of geometrical periodicities. The electrons which passed through S could be further accelerated through a gauze covered hole G into the space charge detector, which was a cylinder D with a pure tungsten filament E 0·1 mm in diameter. The filament was slightly bowed and set at right angles to the axis of the cylinder.

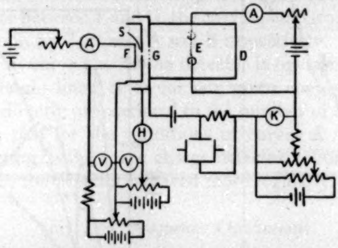


Fig. 1. Circuit and electrodes of experimental tube.

The electrodes other than the filament were made of nickel or nickel–chrome and were enclosed in a soda glass envelope attached through a liquid air trap to a pumping train. The tube was maintained at a temperature of 350°c for 10 hours during pumping, the electrodes were raised intermittently to a red heat by eddy currents, and the filaments were maintained at white heat for half an hour. After this heat treatment mercury was distilled into the tube from a small multi-stage vacuum still into which pure mercury had been previously fractionated. A discharge was passed through the mercury vapour to give the electrodes a final clean up before sealing off the tube, which was done after preheating of the seal.

The circuit used in the experiment is also shown in fig. 1. The cylinder D was made 2 v positive to the negative end of E. The temperature of E was high, and the flow of electrons from E to D was limited by their space charge. The electrons from F were accelerated to S, which, unless otherwise stated, was 8 v positive to the negative end of F. They were then further accelerated through G into D where they collided with gas atoms to produce positive ions. The slow moving positive ions neutralized part of the space charge around E, giving an increased electron current from E to D. This increase in current is normally taken as a measure of the number of ions produced. The main electron current from

E to D, about 2 mA, was balanced out in the galvanometer K by a reverse current produced by a set of fine control potentiometers, and the galvanometer was then used at a sensitivity of 5 mm/μA to measure the variations in current from E to D.

A typical plot of this variation in current with energy of the electrons entering D from F is shown in fig. 2. The operating conditions were: heating current for F, 2·20 A; heating current for E, 1·50 A; potential difference between D and E, 2·00 V; potential difference between S and F, 8·00 V; tube temperature 26°C, and vapour pressure 0·002 mm Hg. No attempt has been made to correct the voltage scale for contact potential difference or for potential differences along the leads. The curve shows irregularities at X and Y which are similar to some of those obtained by McFadden (1946). Further investigation showed however that the voltage spacing of the breaks and their sharpness was highly sensitive to operating conditions. The variation in voltage position of the breaks for different runs was as much as 1·0 V and was probably too large to be caused by varying contact potential differences.

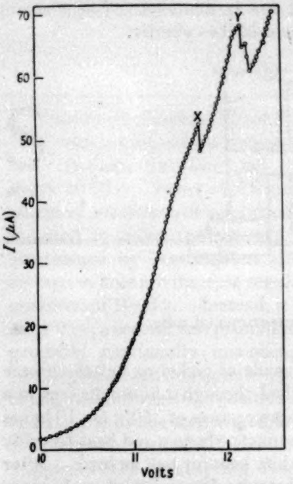


Fig. 2. Variation in the electron current I in the detector as a function of the energy of the ionizing electrons.

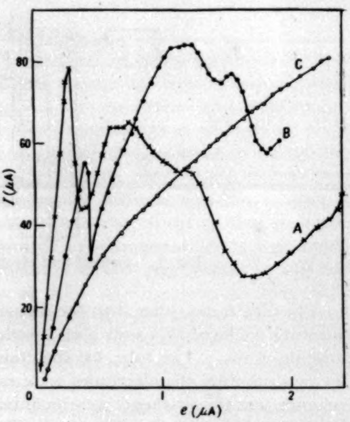


Fig. 3. Response of the space charge detector I as a function of the current of ionizing electrons e for different electron energies.

When recording the variations in current, readings of the galvanometer were taken at intervals of 0·01 V. In many runs it was noticed that the galvanometer became unstable at one reading on the downward break and oscillated with amplitude of about 2 to 3 cm on the scale. For readings of voltage on each side of this the scale reading was stable to 0·5 mm. These measurements indicated that some process was taking place within the tube which was not necessarily connected with ionization probability. To investigate it further the following experiments were performed.

(i) *Velocity Distribution*

The electrons from F were accelerated by 8 v to S and were then retarded by a variable potential difference between S and G. The electron current from F to G was measured as the retarding potential was increased. A graphical differentiation of the decrease in current showed the velocity distribution of the electrons entering G from F to be roughly gaussian with a half width of the order of 0·6 v. It is difficult to understand how the detail shown in fig. 2 could be recorded using this wide velocity distribution if it resulted from atomic processes. In some cases the abrupt changes in current occurred within 0·01 v.

(ii) *Space Charge Detector*

An investigation was made of the changes in current in the space charge detector when the number of electrons entering from F was increased, their energies being kept constant. The intensity of the electron beam entering D from F was altered by varying the potential of S from 0 to 6·0 v positive with respect to F and was measured by the galvanometer H placed in the circuit on the lead to G. The changes in current between E and D were plotted against the intensity of the electron beam entering D from F. For electron energies of 10·8 v and 11·0 v, determined by the potential difference between F and G, the curves pass through a series of maxima and minima as shown in fig. 3, curves A and B respectively. For 12·0 v the curve is smooth, as shown in curve C. If the detector is operating normally, all these curves should be straight lines, since for any given energy the number of ions produced should be directly proportional to the number of the ionizing electrons. The results indicate that for the conditions of curves A and B of fig. 3 some instability was occurring in the space charge detector. For the slightly different conditions of curve C, the detector behaved more regularly.

(iii) *Low Frequency Oscillations*

A test was made in the detector circuit (D to E) for the presence of low frequency oscillations at the discontinuities found in fig. 2. A 1000 ohm resistance was placed in series with the circuit. The variations in the potential difference across the resistance due to the current were amplified and applied to the vertical deflecting plate of a cathode-ray oscillograph. When conditions were adjusted so that the current to D was just about to decrease abruptly, as at X in fig. 2, oscillations were visible on the screen of the oscillograph. They correspond in this case to 20 μA oscillating current when the total change in current between E and D due to the ionization was about 60 μA. The frequency of the oscillation was 300 c/s and the wave form that of a relaxation oscillation. When the voltage between F and G was increased or decreased by 0·2 v the oscillations disappeared.

(iv) *Voltage Position of the Irregularities*

A study of the irregularities showed that they occurred at lower bombarding electron energies (measured by the voltage of G with respect to F) for higher currents of the bombarding beam (dependent on the voltage of S with respect to F), the heating current of F being kept constant. Two curves obtained are shown in fig. 4. For curve A 4 v were applied to S, and the irregularity occurred with 13·20 v on G. For curve B 8 v were applied to S, thus giving a higher current; the irregularity then occurred at 12·25 v. This wandering of the irregularity is a movement with respect to the unvarying initial rise in the curve due to ionization at about 10 v, below which the curves are coincident. It is therefore unlikely to arise from varying contact potential differences, or potential drops along the leads.

The dependence of the voltage position of the irregularity upon the current in the primary electron beam entering D from F was repeatedly studied with similar results by this method or by varying the current through change in the temperature of F. It was noticed in each case that the irregularity occurred at about the same current from E to D in the space charge detector, probably corresponding to a definite ion density.

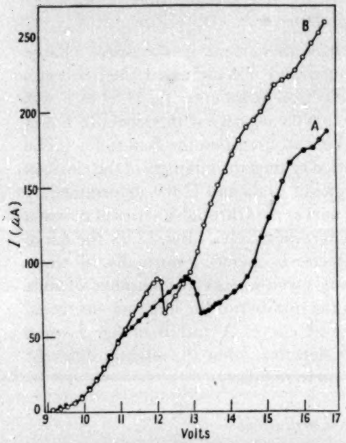


Fig. 4. Variation in the current in the space charge detector I as a function of the voltage of the ionizing electrons, for different potentials on the first grid S.

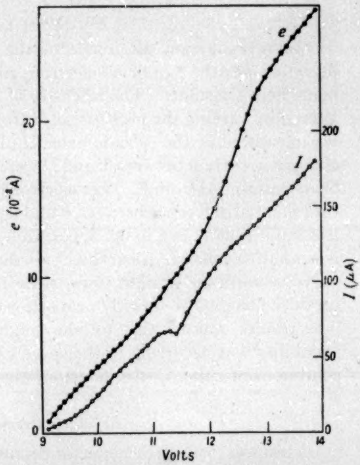


Fig. 5. Ionizing electron current e and space charge detector current I as a function of the voltage of the ionizing electrons.

The voltage position of the irregularity also varied with the heating current of the filament E, and with the anode voltage of D with respect to E. It was lower for greater heating current or for lower anode voltage.

(v) *Primary Electron Current*

The possibility that the irregularities originated in the electron source F and were being carried into the detector by the electron beam was investigated. Figure 5 shows plots of variations in the grid current e, measured by the galvanometer H in fig. 1, and in the detector current I, with electron energy. In this case there is no indication of irregularities in the grid current. Galvanometer sensitivities were such that a change in e produced seven times that change in I. Further, when oscillations were being observed in the detector currents none could be observed in the grid currents. It seems that, in this case at least, the irregularities originate within the detector.

(vi) *Effect of External Circuit*

All experiments were carried out within an earthed metal cage which screened the apparatus from external electrical disturbances. The oscillations were found to occur for various circuit designs. There was no detectable change in frequency when capacities and inductances were introduced into the circuit. We thus conclude that the oscillations do not depend upon the external circuit.

When a 25 mH choke coil was placed in series with the circuit between E and D the oscillations could no longer be observed. However, the characteristic was identical with that taken without the choke, and showed irregularities similar to those in fig. 2. It is concluded that the irregularities in the current characteristic do not depend on the existence of oscillations in the external circuit, and that the associated phenomena of oscillations and irregularities cannot be explained by circuit pick up of electrical disturbances.

§ 3. DISCUSSION

There are three aspects in which the results of these experiments are significant: (i) They indicate that under certain conditions a space charge detector may become non-linear and unstable; (ii) they resolve some of the difficulties associated with the interpretation of the experimental results on the ultra-ionization potentials of mercury vapour by showing that McFadden's results did not necessarily arise from variations in ionization probability; and (iii) they indicate a phenomenon of instability of space charges which, as far as is known, has not previously been reported, and give some basis for a theory of its mechanism.

(i) *Instability of the Space Charge Detector*

Space charge detectors have been used to measure ionization in such experiments as those of Lawrence and Edlefsen (1929) on photo-ionization, Hughes and Van Atta (1930) and McFadden (1946) on ionization in mercury vapour by electron impact, and Varney (1936) on ionization by atomic impacts. It is interesting to note that in each case there were unexplained anomalies. Lawrence and Edlefsen obtained results for the photo-ionization of potassium which were not in agreement with the results of other experiments. Hughes and Van Atta found only changes in slope in the ionization probability curve for mercury vapour, while McFadden reported maxima and minima. Varney's results disagreed with similar work by Rostagni (1934), who used an electrometer directly to measure ionization. It is not possible to make a complete analysis of these experiments but it is suggested that an instability of the space charge detector such as that reported in this work may have been responsible for some of the anomalies.

(ii) *Ionization in Mercury*

The experiments reported here show that McFadden's results cannot be interpreted unambiguously as variations in ionization probability. Similar effects may have been present in the experiments of other workers although it is impossible to be certain because of the differences in the experimental techniques.

(iii) *Mechanism of the Oscillations*

The natural interpretation of the irregularities observed is that they are due to the onset of oscillations. Three considerations indicate however that ion oscillations may not be the sole cause of the effects observed. (*a*) The frequency calculated for ion oscillations under comparable conditions (Kingdon 1923) is not less than 10 kc/s, which is much higher than the frequencies of the oscillations which we have observed. (*b*) The life time of an ion under the experimental conditions is about 10^{-3} sec, which is probably too short to produce oscillations of 300 c/s. (*c*) It is to be expected that an ion in oscillation would have a longer life time than normal. It would thus cause greater neutralization of space charge and a consequent increase in electron current in the detector. Contrary to this it has been found that on the onset of oscillation there is a marked decrease in current.

The low frequency of the oscillations and the fact that they seem to occur when some critical ion density is present within the detector indicates that they are caused by a periodic rearrangement of ion space charge. Consideration of the present experiments in conjunction with others made by Kingdon (1923) shows that the oscillations probably consist of a relatively slow accumulation of ion charge followed by a sudden loss of ions. The loss may be due to a partial disappearance of the potential minimum near the cathode resulting in a more rapid flow of ions to it. This oscillation would repeat itself with a frequency determined mainly by the time required for the ion space charge to accumulate and would cause a corresponding oscillation in electron current to the anode.

§ 4. SUMMARY OF RESULTS

Using a space charge detector by the Hertz method to study the ionization of mercury vapour by electron impact, irregularities have been found in the ionization probability curve above 10·4 v. The irregularities are similar to some reported previously, but because of their variations with tube conditions it is believed that they do not arise from atomic processes and are caused by an instability in the space charge detector. Low frequency oscillations in the current in the detector were found to be associated with the irregularities. It is suggested that this phenomenon may have caused spurious results in other experiments with space charge detectors including experiments on the ultra-ionization potentials of mercury vapour, and that the irregularities and oscillations arise from a discontinuous rearrangement of ion space charge.

ACKNOWLEDGMENTS

The author wishes to thank Professor K. G. Emeléus and Dr. R. H. Sloane for constant assistance and guidance during this work. Some preliminary experiments were performed in co-operation with Dr. D. Greene and Mr. B. Elliot, to whom thanks are also due.

REFERENCES

HUGHES, A. L., and VAN ATTA, C. M., 1930, *Phys. Rev.*, **36,** 214.
KINGDON, K. H., 1923, *Phys. Rev.*, **21,** 408.
LAWRENCE, E. O., 1926, *Phys. Rev.*, **28,** 947.
LAWRENCE, E. O., and EDLEFSEN, N. E., 1929, *Phys. Rev.*, **34,** 1056.
McFADDEN, T., 1946, *Phil. Mag.*, **37,** 630.
NOTTINGHAM, W. B., 1939, *Phys. Rev.*, **55,** 203.
ROSTAGNI, A., 1934, *Nuovo Cim.*, **11,** 34.
STEVENSON, D. P., and HIPPLE, J. A., 1942, *Phys. Rev.*, **62,** 237.
VARNEY, R. N., 1936, *Phys. Rev.*, **50,** 159.

Thermoelectric Effects due to Stationary and Moving Asymmetrical Temperature Gradients in Mercury

By N. FUSCHILLO

The Physical Laboratories, University of Leeds

Communicated by E. C. Stoner; MS. received 17*th March* 1953

Abstract. Stationary and moving asymmetrical and symmetrical temperature distributions are applied to mercury, employing maximum temperatures from the melting point to the boiling point of mercury and an asymmetrical temperature gradient ratio of 1 : 96. No electromotive force greater than 0·005 μv was detected until maximum temperatures greater than 176°c were reached. These e.m.f's were found to be spurious and were traced unambiguously to the finite electrical conductivity of Pyrex glass above this temperature. The experiments confirm the conclusions drawn from earlier work on copper and platinum, and show that the Thomson e.m.f. produced by a temperature difference is independent of the shape of the temperature gradient and the rate at which it is changing.

§ 1. Introduction

In a previous paper (Fuschillo 1952, to be referred to as I) it was shown that the difference in the e.m.f's produced by applying asymmetrical and then symmetrical temperature distributions to specimens of copper and platinum were probably explained by the more extensive 'averaging' of the e.m.f's due to inhomogeneities on the shallow side of the asymmetrical temperature distribution than on the steep side. If the explanation suggested is correct it would be expected that a metallic substance which exhibits no e.m.f. on the application of a symmetrical temperature gradient would also exhibit no e.m.f. on the application of an asymmetrical temperature distribution. Mercury can be expected to be comparatively free from the type of inhomogeneity (strains, inclusions etc.) in which it is envisaged that the symmetrical and asymmetrical effects originate. The experimental methods developed in the previous investigation have therefore been applied to this metal in the expectation that the behaviour would be different from that shown by solid metals. Experiments were performed with maximum temperatures varying from the melting point to the boiling point of mercury in an apparatus producing an asymmetrical gradient ratio of approximately 1 : 96. The specimen was of uniform cross-sectional area, and hence the heat current in the steep temperature gradient was approximately 96 times that along the less steep gradient, thus producing a net flow of heat through the conductor. This experiment, therefore, also provides a means of examining whether the irreversible heat flow due to thermal conduction, which necessarily accompanies the production of the Seebeck e.m.f. in a thermoelectric circuit, itself contributes to this e.m.f. (Giauque, Buffington and Shulze 1927).

In the treatment of the thermoelectric circuit given by Bruzs (1935) an e.m.f. is predicted due to the application of a temperature gradient under transient conditions. An attempt to detect this e.m.f. was made by rapidly moving an asymmetrical temperature distribution of ratio 1 : 96 along the mercury specimen. Finally, to ensure that all the electrical effects which might be produced by an asymmetrical temperature distribution were investigated, an electric current was reversed through the asymmetrical heat flow under both equilibrium and transitory conditions. Measurements on any change in the magnitude of this current on reversal were made to verify that no interaction occurred between the irreversible heat current and the electric current.

§ 2. Previous Work

The results obtained with mercury in previous attempts to measure asymmetrical temperature gradient e.m.f's show little agreement. Electromotive forces have been measured by Benedicks (1917, 1919), Gouineau (1920), Tsutsui (1929) and by Ch'en and Band (1936). On the other hand Haga and Zernike (1919) obtained a maximum e.m.f. of only $0{\cdot}05\,\mu$v, while Benade (1921) and Fraser (1938) could not detect an e.m.f. greater than $0{\cdot}01\,\mu$v. Over half of these workers employed the constriction method, the others asymmetrically heated mercury contained in glass tubes of uniform diameter with non-uniformly wound electric furnaces. The first method is open to the following objection. The production of an asymmetrical temperature gradient e.m.f. may be dependent upon an asymmetrical flow of heat in the specimen. The steep temperature gradient is produced by constricting the specimen and heating on one side of the constriction. This procedure, however, severely limits the asymmetry of the heat flow as the thermal current through a cross section of a conductor is directly proportional to both the cross-sectional area and the temperature gradient. With the second method only small (i.e. less than 1 : 7) asymmetrical gradient ratios can be applied. The limitations of both of these methods can be overcome by the present procedure, in which an asymmetrical gradient ratio of 1 : 96 is applied to a specimen of uniform cross-sectional area.

§ 3. Present Work

The general method of investigation was similar to that described in I. Platinum wires, in the tubes a and a′ of fig. 1, were used to lead the e.m.f. developed in the mercury system to the galvanometer, and platinum–iron thermocouples were used, in a manner described later, to verify that the mercury platinum junctions were at equal temperatures.

(i) *Experimental Details*

Vacuum distilled mercury, outgassed and dried by pumping over it for six hours at a temperature of 100°c, was used to fill the apparatus illustrated in fig. 1. Here A is a brass former of thickness 1·5 mm on which was wound a coil of Pyrex tubing of axial length 41 cm and diameter 11·5 cm, with zero spacing between successive turns. The Pyrex tubing had an internal diameter of 2·0 mm and an external diameter of 3·0 mm. The system had been cleaned with chromic acid and distilled water, after which it was thoroughly dried. The mercury in the reservoirs D and D′ could be circulated through the coil by opening taps E and E′, the tubes C and C′ acting as air traps. A length of 40 cm of the Pyrex–mercury leads from C and C′ was immersed in the mercury H contained in a Pyrex tube, which was surrounded by oil G contained in a Dewar vessel. The leads terminated in the sealed tubes a and a′ each of which contained a junction (b or b′ of fig. 2) of one platinum and two iron wires, the junction only being in electrical contact with mercury. To the other ends of these wires were soldered copper leads connected to thermoelectrically neutral reversing switches 1 to 5 (fig. 2). The junctions cc′, dd′, ee′ of these wires were insulated with ‘Sleek’, wrapped together and placed between the tubes a and a′. Thus immersed in the mercury H these junctions were very nearly at the same temperature as b and b′. With the aid of switch 1 and galvanometer G_1 the e.m.f. developed in the mercury could be measured, and with G_1 and switches 2 and 3 the equality in the temperature of the junctions b and b′ could be ascertained with the platinum–iron thermocouples by the method outlined for platinum in I. With the arrangements described no difficulty was experienced in maintaining b and b′ at the same temperature to 0·001°c, which enabled measurements significant to 0·005 μv to be made. The galvanometer employed was the same as that used in the previous experiments with copper and platinum. The resistance boxes A, B and C in conjunction with an accumulator and switches 4 and 5 enabled an electric current to be reversed through the mercury circuit. This current was measured with the galvanometer G_2.

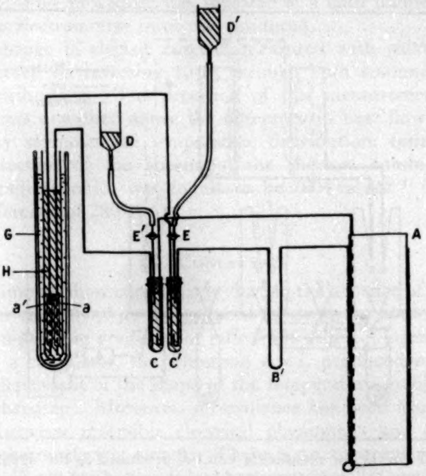


Fig. 1. Apparatus for producing stationary or moving asymmetrical temperature gradients in mercury contained in Pyrex tubing. Asymmetrical gradient ratio 1 : 96.

(ii) *Precautions against Spurious Effects*

As a safeguard against electrical interference from outside sources the apparatus was electrically shielded and the oil heaters switched off while a measurement was taken. The test for electrical leakage between the coil and the former, described in I, was performed at regular intervals. In addition the galvanometer G_1 could be included in a circuit to measure the magnitude of such a leakage current. The mercury in a and a′ of fig. 1 was occasionally subjected to a pressure of approximately one metre of mercury. If leakage of mercury occurred through the rubber and black wax sealing these tubes, the mercury coil could have been short circuited. Hence, the insulation resistance between the junction b of fig. 2 and the mercury H was frequently checked. The resistance of the mercury circuit was measured at every reading to ensure that the electrical connection was complete. The simple electrical circuits associated with these arrangements are not shown in fig. 2.

(iii) *Experimental Procedure and Results*

Asymmetrical temperature gradients were produced by immersing the first six turns of the coil in electrically heated oil for maximum temperatures above 20°c, or acetone cooled with liquid nitrogen for lower temperatures. Symmetrical gradients were similarly applied by immersing the loop B′, illustrated in fig. 1.

No symmetrical gradient e.m.f. could be detected for maximum temperatures varying from the melting point to the boiling point of mercury. When the mercury solidified, however, electromotive forces were produced. An e.m.f. of 0·12 μv was recorded when the acetone bath was at a temperature of −41°c. This effect was undoubtedly caused by the inhomogeneous solidification of the mercury. No e.m.f. was observed on applying asymmetrical temperature gradients until a temperature above 176°c was reached, but then the insulation resistance between the coil and the former A was no longer infinite. The e.m.f. increased and the insulation resistance decreased with increasing maximum temperature. The results obtained at temperatures in excess of 176°c showed little uniformity, the electromotive force produced depending on the thermal history of the run. A typical example of this effect is as follows.

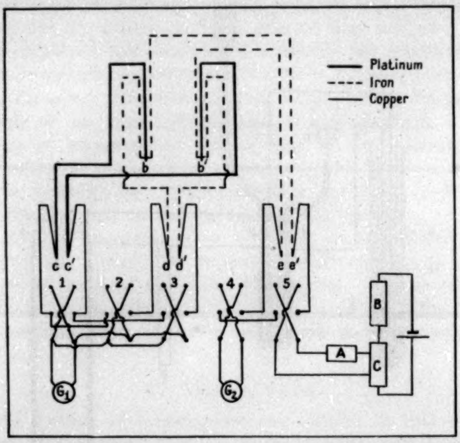


Fig. 2. Electrical circuit for measuring e.m.f's, produced in the apparatus illustrated in fig. 1. Junctions b and b′ are in electrical contact with the mercury contained by tubes a and a′ of fig. 1.

$T_{\max}$ (°c)	176	185	230	300
Asymmetrical gradient e.m.f. (μv)	0·00	0·06	0·14	0·21
Insulation resistance (ohms)	∞	$5·0\times10^6$	$8·5\times10^5$	$9·0\times10^4$
Leakage current (μA)	0·00	0·12	0·64	0·83

The electromotive forces observed were undoubtedly due to the significant electrical conductivity of the Pyrex glass at the higher temperatures, while the rather large values of the leakage current were probably caused by the e.m.f. developed in the mercury–Pyrex–brass system.

Transitory conditions were produced by opening the taps E and E′ of fig. 1, thus allowing mercury to circulate through the coil. Experiments similar to those already described were then repeated at a bath temperature of 150°c. The rate of flow was nearly 50 $\mathrm{cm}^3\,\mathrm{min}^{-1}$. Thus an asymmetrical temperature distribution, with an asymmetrical gradient ratio of approximately 1 : 96, was moved along a mercury conductor at the rate of 26 $\mathrm{cm\,sec}^{-1}$. No e.m.f. was observed during the circulation of the mercury or on immediately starting or stopping the flow. The experiment, therefore, failed to verify the prediction of Bruzs (1935). This procedure was repeated at a bath temperature of 220°c but no change in electromotive force was produced.

Finally no change in electric current, measured with galvanometer G_2 of fig. 2, was observed on reversing 10 μA through both stationary and moving temperature distributions. The precision of this measurement was 0·01 μA. In the experiments described above the difference in heat flow along the sides of the stationary asymmetrical temperature distribution, estimated from the area of cross section of the specimen, the thermal conductivity and the asymmetrical gradient ratio, was found to be 0·04 $\mathrm{cal\,sec}^{-1}$ for a maximum temperature difference of 280°c.

§ 4. CONCLUSION

These experiments show conclusively that, in the absence of spurious effects, no e.m.f. greater than 0·005 μv is produced by applying stationary or moving asymmetrical temperature gradients of ratio 1 : 96 to a homogeneous conductor. Hence, in such a conductor, the Thomson e.m.f. produced by a temperature difference is independent of the shape of the temperature gradient and the rate at which it is changing. Moreover, no evidence has been found for any kind of interaction between reversible electrical phenomena and irreversible heat transfer in the thermoelectric circuit. There is no apparent reason why these conclusions should not apply to any homogeneous metallic conductor.

ACKNOWLEDGMENTS

I am indebted to Professor E. C. Stoner and Dr. F. E. Hoare for their interest and helpful discussion.

REFERENCES

BENADE, J. M., 1921, *Phys. Rev.*, **18,** 199.

BENEDICKS, C., 1917, *C. R. Acad. Sci., Paris*, **165,** 426; 1919, *Ibid.*, **169,** 578.

BRUZS, B., 1935, *Proc. Roy. Soc.* A, **151,** 640.

CH'EN, J. L., and BAND, W., 1936, *Proc. Phys. Soc.*, **48,** 164.

FRASER, M., 1938, *Phil. Mag.*, **25,** 785.

FUSCHILLO, N., 1952, *Proc. Phys. Soc.* B, **65,** 896.

GIAUQUE, W. F., BUFFINGTON, R. M., and SHULZE, W. A., 1927, *J. Amer. Chem. Soc.*, **49,** 2343.

GOUINEAU, M., 1920, *C. R. Acad. Sci., Paris*, **170,** 1567.

HAGA, H., and ZERNIKE, F., 1919, *Proc. K. Ned. Akad. Wet.*, **21,** 1262.

TSUTSUI, T., 1929, *Inst. Phys. Chem. Research, Tokyo*, **8,** 19.

The Design and Operation of a 4·5 MeV Microtron

By C. HENDERSON, F. F. HEYMANN and R. E. JENNINGS
University College, London

Communicated by H. S. W. Massey; MS. received 16*th February* 1953

Abstract. The design and construction of a 4·5 MeV microtron is described. The machine operates at a wavelength of 10 cm, at a magnetic field of about 1000 gauss. The diameter of the final orbit is approximately 30 cm. A circulating current of about 0·2 μA mean has been observed at a duty cycle of 4×10^{-4}, with a 500 kW peak microwave source.

A simple and efficient method of beam extraction has been developed. This has enabled a beam of semi-angles 1·5° in the horizontal plane and 0·3° in the vertical plane to be brought into the laboratory.

It is concluded that with sufficient radio-frequency power, the output of the microtron should compare favourably with that obtainable from linear accelerators.

---

§ 1. INTRODUCTION

ALTHOUGH the principle of operation of the microtron was established at the same time as that of the synchrotron (Veksler 1945, MacMillan 1945), the former instrument has received much less attention than the latter. This reluctance to investigate the practical possibilities of the microtron seems to be based largely on the lack of positive magnetic focusing of the particles and on the difficulties associated with injection.

The first published account of the operational characteristics of a microtron is that of a machine constructed by Redhead, LeCaine and Henderson (1950). Judging by their experience, and that of the present authors, the lack of magnetic focusing is not serious; the electric focusing due to the resonator field is apparently enough to prevent a serious loss of particles. The problem of injection was overcome by accelerating electrons from rest, using field emission from the resonator lips as a source of electrons.

When a high energy electron scattering programme was started in these laboratories, it was felt that two important features of a suitable accelerator should be (*a*) ease of extraction of the electron beam, (*b*) ease of providing a beam of constant and known energy. The first of these is met by the microtron, in so far as the orbits are well spaced, which makes it possible to provide a magnetic-field free path from the final orbit without interfering appreciably with the field at the preceding orbit. The second condition is satisfied, since the microtron requires a steady field, as distinct from the alternating magnetic fields of both the betatron and electron synchrotron. This makes it possible to stabilize the field with ease and the output energy is determined by the value of the magnetic field and the position of the extractor mechanism.

For these reasons, it was decided to construct a small microtron, in the first instance to act as a source of electrons for scattering experiments at 4·5 MeV, and secondly to provide practical information about the behaviour of this type of accelerator, as a guide to the design of a larger machine.

The microtron has been found to be remarkably stable in operation, in spite of the fact that field emission is relied upon to provide the electrons to be accelerated, and the beam has been brought out of the machine by very simple means. With the confidence gained during the operation of this small machine, the design of a 30 MeV microtron has been commenced.

The principle of magnetic resonance on which the operation of the microtron rests has been discussed in greater detail by various authors since its introduction by Veksler (Redhead *et al.* 1950, Itoh and Kobayashi 1950, Henderson *et al.* 1953).

The accelerator operates in the manner described below.

A microwave resonant cavity, the shape of which is illustrated in fig. 4, is placed near the edge of a steady, uniform magnetic field, with its axis of symmetry perpendicular to the direction of the magnetic field. The resonator is excited so that the peak voltage across the lips is slightly larger than the voltage corresponding to the rest mass of the electron (about 0·511 MeV). Electrons are emitted from one of the lips by field emission and those which cross the gap at the appropriate phase of the electric field emerge from the hole in the opposite cone of the cavity with a total energy of two rest masses. The values of the magnetic field and the operating frequency of the cavity are adjusted so that such electrons require a time corresponding to two cycles of the radio-frequency field to complete their first orbit. They will then make the next transit of the cavity at the correct phase for each electron to gain one additional rest mass of energy. Since the time needed for an electron to complete an orbit in the magnetic field is directly proportional to its total energy, the second orbit corresponds to a time interval of 3 cycles of the radio-frequency field and the electrons under consideration once more arrive at the cavity at the appropriate phase for each to receive one rest mass of energy. In this manner the electron is subjected to continued acceleration, moving on orbits of increasing radii, all cotangential to the axis of the resonator. Thus, after n transits of the accelerating field, the total energy of an electron is $n+1$ rest masses.

Other modes of operation are possible. If one considers electrons starting from rest, it can readily be shown that the possible conditions for resonance are given by (Redhead *et al.*)

$$B=\omega \mathbf{c} m_0/\mathbf{e}(a-b) \qquad \text{(gaussian units)} \qquad \ldots\ldots(1)$$

$$V=\frac{m_0\mathbf{c}^2}{\mathbf{e}}\,\frac{b}{a-b} \qquad \text{(gaussian units)} \qquad \ldots\ldots(2)$$

where B is the magnetic flux density, V the peak potential difference across the cavity, ω the angular frequency of radio-frequency supply, m_0 the electron rest mass, and a and b any integers such that $1\leqslant b<a\geqslant 2$. The mode described above is the fundamental, corresponding to $a=2$, $b=1$. As there is no mode requiring a higher magnetic field, this is one of a set (corresponding to $a-b=1$) which requires the smallest possible magnet radius for a given final energy. Moreover, the mode of this set with $b=1$ requires the lowest voltage of the set.

At higher frequencies lack of a sufficiently powerful radio-frequency source makes it imperative to operate in one of the higher modes, requiring a lower value of the voltage. A 1 MeV microtron, operating at 3 cm in the mode

$a=4$, $b=1$, has recently been constructed in the United States (Kaiser 1952). This seems to be about the best that can be achieved at a wavelength of 3 cm with present-day magnetrons.

The principle of acceleration outlined here possesses a phase stability broadly analogous to that of a synchrotron, by virtue of which an electron which crosses the accelerating gap with errors in energy and in phase undergoes a motion in the following orbit which tends to correct these errors. Various aspects of the limits of error in energy and phase within which phase stable acceleration takes place have been investigated (Itoh and Kobayashi, Henderson *et al.*, Redhead *et al.*, Veksler) and agreement of these calculations with the operational behaviour of the machine will be discussed in §6.

§2. MAGNET

The microtron described here (see Plate I) is similar to that constructed by Redhead *et al.* It operates at a frequency of 3030 Mc/s in the mode $a=2$, $b=1$, so that the magnetic flux density required in the air-gap is 1070 gauss. This relatively low value has permitted the construction, by Mr. A. J. Tyrrell of Mullards, of a magnet of novel appearance, the major dimensions of which are illustrated in fig. 1. The pole pieces A are mounted on square steel plates B, spaced apart by four corner pillars C. The plates B and pillars C form the magnetic return path for the flux across the air-gap. Each of the exciting coils D consists of 1300 turns of No. 14 s.w.g. cotton covered copper wire. The two coils are connected in series. The solid brass rings E, of 14 cm × 0·5 cm cross section, act as damping resistances to limit the inverse voltage surge across the winding when the magnet is disconnected from the supply. As a further protection, a metal rectifier is connected across the windings with reversed polarity.

The magnet requires approximately 4·5 amperes at 100 volts for full excitation. This power is obtained from 3 phase, 400 volt a.c. mains, through six grid controlled mercury vapour rectifiers. The magnet current is stabilized by balancing the voltage drop across a resistor in series with the magnet against the e.m.f. of a cell by a galvanometer and photocell unit. The amplified error signal is used to vary the firing time of the thyratrons in such a sense as to reduce the difference between the voltage across the resistor and that of the cell. By this means, the magnet current can be held constant to one part in a thousand for several hours, despite normal fluctuations in the supply voltage and a 15% increase in the winding resistance due to the temperature increase of the winding during operation. An early tendency towards hunting in the stabilizing system was cured by the addition of a resistance in series with the magnet, in order to lower the effective Q of the magnet.

The absolute calibration of the magnetic field was made with a nuclear resonance circuit. The relative radial variation of the field was determined with a pair of flip coils, using a null method for comparing the field at any desired radius with the field at the centre of the magnet. The distribution obtained in the case of flat pole faces is shown in fig. 2, curve *a*. Subsequently the pole faces were shimmed in order to obtain a larger region of substantially uniform field. The tightly fitting shimming rings are secured by means of grub screws pressing into the cylindrical surfaces of the original pole pieces. The radial variation of the field after shimming is shown in fig. 2, curve *b*. If the machine is operated with the centre of the resonator in a field 1% below the central value, the defocusing region of curve *b* is not serious and the use of the shims enables the maximum output to be increased from 4 to 6 MeV.

The upper pole can be lifted off the magnet and moved laterally by means of a small travelling crane in order to provide access to the vacuum chamber, situated in the air-gap.

§ 3. Vacuum System

The vacuum tank is constructed of brass with an internal diameter of 21 in., an internal height of $3\frac{3}{4}$ in., and a wall thickness of $\frac{3}{8}$ in. Flanged ports are provided as illustrated in fig. 3, in order to accommodate the vacuum pumping manifold, a Philips type vacuum gauge, the waveguide feeding the resonator, the beam extractor and a movable port for investigating individual electron orbits. Two additional viewing ports with glass windows are provided. All seals to the flanges are made with the aid of rubber 'O' rings. The lid of the tank is removable and is sealed to the body by a double 'O' ring seal on a grooved flange on the tank body, with provision for pumping out the region between the two seals. In practice it has never been found necessary to make use of this facility.

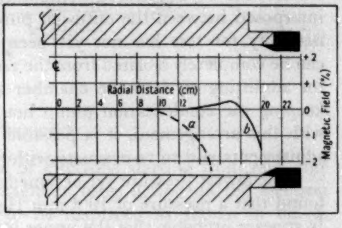


Fig. 2. Magnetic field distribution in the median plane.

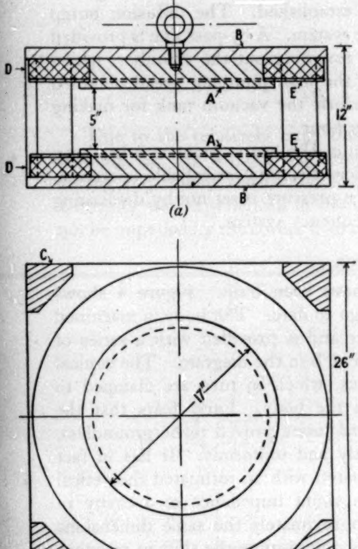


Fig. 1. Magnet. (*a*) Section. (*b*) Plan.

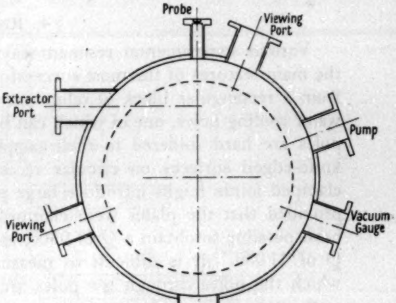


Fig. 3. Vacuum tank—plan view.

The probe is situated diametrically opposite to the resonator, and can be moved along this diameter by a remotely controlled motor operating on the probe shaft via a rack and pinion. The probe shaft moves through a spring loaded Wilson seal. The probe position is indicated at the control panel with the aid of a slide-wire potentiometer, the jockey of which is mounted on the probe shaft.

The whole radio-frequency transmission system, up to the magnetron input unit, which itself operates at atmospheric pressure, consists of an evacuated waveguide. Various tuning and matching adjustments are made through sliding seals.

The vacuum system consists of an Edwards 4 in. diffusion pump, charged with Dow Corning 702 silicone oil, backed by an Edwards 1S–150 rotary pump fitted with a phosphorus pentoxide vapour trap. A large liquid air trap is interposed between the diffusion pump and the vacuum chamber, although the necessity for this has not yet been clearly established. The diffusion pump can be completely isolated from the rest of the system. A by-pass line is provided for admitting air into the chamber and for roughing out the chamber, whilst keeping the condensation pump hot. With the high pumping speed obtained with this arrangement, it is possible to dismantle the vacuum tank for making adjustments and to re-evacuate within half an hour.

The lowest pressure yet obtained is 10^{-5} mm Hg, but in practice it has been found that a pressure of 10^{-4} mm Hg is sufficiently good for reliable operation. It appears probable that the upper limit to the pressure is set not by defocusing of the beam, but by flashover in the radio-frequency system.

§ 4. RESONATOR

Various experimental resonant cavities have been built. Figure 4 shows the main features of the most successful design to date. The body is machined from a rectangular block of tellurium copper and is provided with a series of water cooling holes, one of which can be seen at B in the diagram. The conical poles are hard soldered to their copper discs, which in turn are clamped to knife-edged surfaces on circular recesses in the body. Early fears that the clamped joints might introduce large electrical losses proved to be groundless, provided that the plates were clamped tightly and uniformly. It has in fact been possible to obtain a Q of 9500, as compared with an estimated theoretical Q of 11 000. It is difficult to measure the shunt impedance of a cavity in which the holes through the poles are of approximately the same dimensions as the gap spacing. An estimate was made by determining the shift in resonant frequency when a dielectric rod was inserted along the axis of the cavity. With the assumption that the gap is effectively increased at each end by an amount equal to the radius of the hole, this measurement gave a value of 1·2 megohms.

The cavity block is soft soldered directly to the waveguide, and electrical coupling is provided between the cavity and the waveguide by a hole D in the block, as shown in fig. 4. The diameter of this hole was adjusted experimentally to give the desired coupling.

Frequency tuning of the cavity is accomplished by means of a thermal expansion device which moves the cone A axially relative to the body. A corrugation, not shown in the drawing, was made on the copper disc on which this cone is located in order to facilitate movement of the cone. This mechanism is capable of tuning the resonant frequency through the range 3030 ± 5 Mc/s. This method of tuning is very simple and is practically free from backlash, but has the disadvantage of a thermal delay of about 30 seconds. However, in practice the radio-frequency system is sufficiently stable to make retuning of the cavity seldom necessary. A mechanical tuner operating through a Wilson seal has been used successfully on another model.

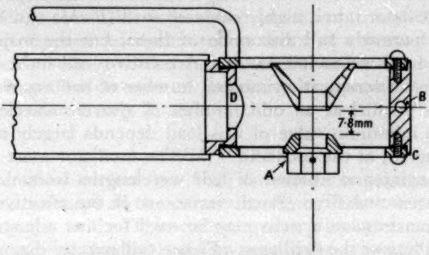


Fig. 4. Cavity resonator.

Due to the relatively large transit time of electrons starting from rest and crossing the gap for the first time, the magnetic force acting on the electron during the transit causes the first orbit to be displaced as shown in fig. 5. Difficulty was experienced in designing a cavity for which the first orbit would not be impeded by the corner C of the cavity. The problem was eventually solved

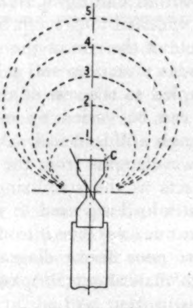


Fig. 5. Displaced first orbit.

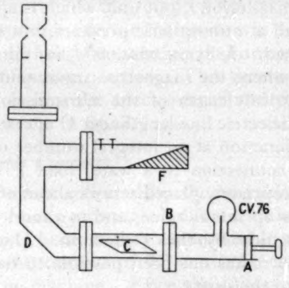


Fig. 6. Waveguide system.

by making the cones asymmetrical, as shown in fig. 4. With this arrangement, the first orbit is placed symmetrically with respect to the external dimensions of the cavity, when the initial transit takes place from the long towards the short cone. The degree of asymmetry required was estimated by a numerical calculation of typical initial transits, assuming a uniform electric field in the gap.

Field emission from the cavity lips is used to yield the electrons to be accelerated, this emission being controlled by first polishing and gold plating the cones, to keep the normal field emission to a low level, and then inserting in the long cone a cylindrical phosphor bronze sleeve of $\frac{1}{64}$ in. wall thickness. By adjusting the distance that this sleeve protrudes into the gap, it is found that the emission can be readily controlled. This system has the advantage of enhancing the emission in the desired direction. The inside of the cavity can be observed directly through a small window shown at D in fig. 6.

§ 5. RADIO-FREQUENCY SYSTEM

Certain conditions have to be fulfilled in order to feed power stably from a self-excited oscillator into a highly resonant load (Pierce and Shephard 1947). Space does not permit a full discussion of these, but the major requirements in practice may be listed as follows: (*a*) A resistive load must be connected to the supply, either in series, at an integral number of half wavelengths from the resonator, or in parallel at an odd number of quarter wavelengths from the resonator. The minimum value of this load depends largely on the character of the Rieke diagram of the oscillator. (*b*) The oscillator must be effectively at approximately an integral number of half wavelengths from the resonator, for optimum frequency stability. Small variations of the effective length of this section of the transmission system may be used for fine adjustment of tuning, by frequency-pulling of the oscillator. Figure 6 illustrates diagrammatically the way in which these requirements are met.

Standard 3 in. × $1\frac{1}{2}$ in. brass waveguide is used throughout in the transmission system. A type CV. 1480 (selected CV. 76) pulsed magnetron operating at 3030 Mc/s is the source of radio-frequency power. The magnetron works at 500 kw peak power output with a nominal pulse duration of $2\,\mu$sec at a repetition frequency of 200 pulses per second. The coupling of the magnetron to the transmission system is adjusted by controlling the depth of insertion of the output probe and by means of a plunger, as shown at A in fig. 6.

The magnetron input unit which is specially widened to prevent breakdown is operated at atmospheric pressure, but the whole of the rest of the waveguide is evacuated. A Pyrex window $\frac{3}{32}$ in. thick provides a vacuum seal at the choke flange B where the magnetron input unit is coupled to the rest of the system. The electrical length of the transmission line can be varied by means of a tapered dielectric line lengthener C operated through a Wilson seal. An E-plane series T-junction at an integral number of half wavelengths from the resonator provides connection to a water load F which acts as the stabilizing resistor Two appropriately placed screws ahead of the water load are used to adjust the load to its optimum value, and to cancel the shunt susceptance introduced into the waveguide by the T-junction. Due to the poor Rieke diagram of the magnetron it has not been possible to feed more than about 50% of the total power into the cavity.

The cavity voltage is monitored with the aid of a bolometer, coupled to a series window which is placed in the waveguide at an odd number of quarter wavelengths from the cavity.

§ 6. OPERATIONAL CHARACTERISTICS

(i) *Resonator Voltage*

When the machine is adjusted for maximum current in the eighth orbit, it has been observed that operation takes place at a magnetic field of 1030 gauss instead of the 1070 gauss required by the elementary theory of the microtron (with injection from rest). It is thus apparent that under optimum conditions there is not sufficient radio-frequency power available to excite the cavity fully to yield 511 kev gain per transit. Operation takes place, instead, at 490 kev gain per transit. The fact that a well-defined set of orbits is obtained at all under these conditions lends strength to the argument advanced by the authors (Henderson *et al.* 1953) that with injection from rest, resonance may be obtained when the voltage differs by a small fraction from 511 kv, if the magnetic field is different from the theoretical value by the same fraction.

The peak voltage obtainable is less than one is led to expect on the basis of power input and the estimated shunt impedance of the cavity. This is not surprising, as steady state conditions are never reached in the cavity during the pulse: the build-up time constant of the cavity is approximately 1 μsec, whilst the total pulse duration is only 2 μsec.

(ii) *Emission*

The total emission current has been collected on a large insulated brass septum placed to intercept all the current emerging from one side of the cavity using a low magnetic field. This current pulse, displayed on a fast triggered oscilloscope, has verified that the build-up time of the resonator is an appreciable fraction of the total pulse duration and that the cavity voltage is only large enough to cause field emission during the latter half of the radio-frequency pulse.

The septum was also used to determine whether the polishing and plating of the cavity were effective in limiting the emission from the cone facing towards the backward direction. This was done by comparing the current collected by the septum with the magnetic field in the normal direction with that collected with the magnetic field reversed. By this means, the wanted and unwanted currents were swung on to the septum in turn. The reverse current was found to be only about one twentieth of the current in the required direction.

The energy spectrum of the electrons emitted from the cavity was determined by operating the machine at a low magnetic field and recording the current collected by the movable probe as a function of probe position. This measurement in effect utilizes the magnet as a 180° focusing flat spectrometer and the energy spectrum obtained in this manner is shown in fig. 7.

(iii) *Orbit Current*

Figure 8 is a typical plot of the current collected by the movable probe as a function of probe position under normal operating conditions. The last two orbits are of decreased magnitude largely because they are located in the region of non-uniform magnetic field. A mean current of 0·2 μA in the eighth orbit is typical of the best performance of the machine. Since the voltage across the cavity is high enough for resonance only during the latter half of the pulse, this mean current corresponds to a pulse current of over 1 mA. Although only ten orbits are shown in fig. 8, eleven orbits have been observed under favourable conditions, with 0·08 μA in the eleventh orbit. The machine was originally designed for only eight orbits but as electronic loading due to later orbits is small, the final energy can be increased by using a larger magnet.

The orbits have been examined visually by means of a fluorescent material deposited on the movable probe. The first orbit was found to be rather badly focused, being about 3 cm high and 4 cm wide, although the width must be accounted for largely in terms of the energy spread of the emitted electrons.

The other orbits all have a vertical spread of about 5 mm. As this is somewhat smaller than the diameter of the orifice of the cavity, it is apparent that no appreciable defocusing of the beam takes place. This is also obvious from the fact that the currents in all the orbits beyond the third are approximately constant, until the non-uniform region of the magnetic field is reached.

The microtron has a mode of operation at half the normal magnetic flux density for which the voltage gain per transit has the normal value, corresponding to $a=4$, $b=2$ in eqns. (1) and (2). Orbits of this mode have been observed, at currents much lower than those obtained in the fundamental mode.

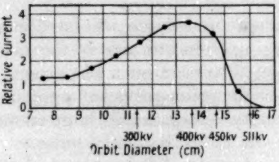


Fig. 7. First orbit spectrum.

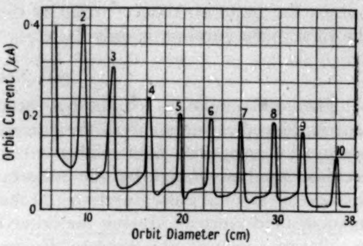


Fig. 8. Current in various orbits.

(iv) *Variation of Current with Magnetic Field*

The output current was found to be more critically dependent on the value of the magnetic field (fig. 9) than that reported by Redhead *et al.* (1950), but agrees fairly well with the calculated value of the limits of phase stability. It has been shown (Henderson *et al.* 1953) that for this microtron, phase stable acceleration is possible with the normal value of the magnetic field only for peak cavity voltages between 520 kv and 614 kv (taking into account the finite transit time), corresponding to a fractional range of 1·18 : 1 in voltage. In the same paper, it was shown that a small fractional change in magnetic field is equivalent to the same fractional change in voltage, so that, at constant cavity

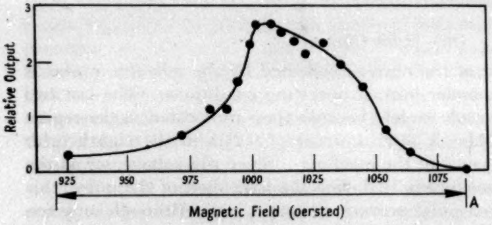


Fig. 9. Variation of output with magnetic field.

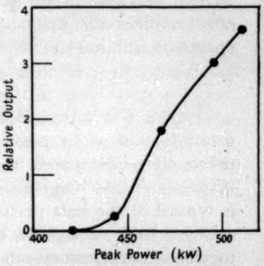


Fig. 10. Variation of output with magnetron power.

voltage, one would expect acceleration to occur over a range of 1·18 : 1 in magnetic field. This range is shown in fig. 9, with the upper limit taken arbitrarily at the point A, from which it is apparent that the range over which appreciable output is obtained agrees reasonably with the expected range.

(v) *Variation of Current with Magnetron Output Power*

The current in the eighth orbit is plotted as a function of magnetron output power in fig. 10. From the steep slope of this curve it is fair to expect that an increase in radio-frequency power will result in a large increase in the output of the machine. Unfortunately, both the magnetron and its modulator are already overloaded at present, and all attempts to increase the magnetron power beyond that shown in fig. 9 have been unsuccessful.

The present output of the machine is, however, ample for the experimental programme envisaged.

(vi) *Beam Extraction*

Since two adjacent orbits in the fundamental mode require intervals of n and $n+1$ cycles of the radio-frequency respectively for completion, then for orbits where the electron velocity is near to that of light, the increment in path length between orbits becomes sensibly equal to the wavelength of the radio-frequency supply. At 10 cm wavelength, this corresponds to a spacing of 3·18 cm between orbits at a position diametrically opposite the cavity. With this spacing the extraction of the beam becomes a simple matter. All that is required, in principle, is to insert a tube of suitable magnetic material into the vacuum chamber so that the axis of the tube intercepts the desired orbit tangentially, when the beam should proceed straight down the tube, due to the production of a field-free region inside the tube. The tube should be of small enough diameter to cause no appreciable interference with the magnetic field at the previous orbit. Such an extractor has been constructed. The tube is made of mild steel, tapered from $\frac{3}{4}$ in. i.d. × 1 in. o.d. at the edge of the magnet to $\frac{1}{4}$ in. i.d. × $\frac{13}{32}$ in. o.d. at the point where the orbit is intercepted. About 50% of the circulating current has been brought out of the ninth orbit into the laboratory through a 0·004 in. aluminium window.

The divergence of the beam has been measured in a vacuum and semi-angles were found to be 1·5° in the horizontal plane and 0·3° in the vertical plane. Plate II is a photograph of the extractor in position.

(vii) *Measurement of Particle Energy*

The kinetic energy of the particles as calculated from the magnetic field and orbit radius in the machine give a value of 4·5 MeV with an upper limit to the energy spread (estimated from the cross section of the final orbit and the size of hole in the cavity) of ± 0·07 MeV.

An independent measurement of the energy of the electron beam was made by determining its range in aluminium. Two ionization chambers, one on each side of the absorber, were used and the current in them compared directly by means of a potentiometer. Making use of the range energy data given by Katz and Penfold (1952) the value found for the 'practical range' corresponds to a kinetic energy of 4·6 MeV.

§ 7. CONCLUSION

Although the output of the microtron is more than enough for experimental purposes it is evident that the radio-frequency power at present available is insufficient to allow us to make a fair comparison with linear accelerators. It is, however, expected that with an input power in the region of a megawatt, the microtron will compare favourably with linear accelerators. This is justified by the fact that any increase in power fed to the cavity can be used for electronic loading, no additional power being required for excitation. The steep rate of rise of power with magnetron power (fig. 10) lends strength to this argument. The possibility of using a 2 MW magnetron for the sake of comparison with linear accelerators is at present being explored.

The output has been found to be unexpectedly stable, in view of the fact that the machine relies on field emission to provide a source of electrons. It was, for instance, possible to retrace curves such as that shown in fig. 8 point by point over a period of at least an hour, yielding substantially a constant curve. This is all the more remarkable in view of the small margin of power (fig. 10) available for producing a beam.

In view of the simplicity of the machine and the small amount of precision engineering involved compared with a linear accelerator, the authors feel that the microtron offers a competitive alternative in the energy region below, say, 50 MEV. The relatively large final orbit diameter compared with that in a synchrotron or betatron need not be taken as a serious disadvantage, since (*a*) the low flux density involved results in a magnet of comparatively low weight, and (*b*) the magnet is much simpler to construct since the microtron operates with a constant, as opposed to an alternating, magnetic field.

ACKNOWLEDGMENTS

In conclusion, the authors wish to thank Professor H. S. W. Massey for giving them the opportunity of developing the machine and for his continued interest in the project, to Dr. E. H. S. Burhop for his valuable advice on many points, and to those members of the technical staff of the Physics Department who helped so willingly with the construction.

REFERENCES

HENDERSON, C., HEYMANN, F. F., and JENNINGS, R. E., 1953, *Proc. Phys. Soc.* B, **66,** 41.
ITOH, J., and KOBAYASHI, D., 1950, *Col. Fac. Sci.*, Osaka University, B, **11.**
KAISER, H. F., 1952, *Bull. Amer. Phys. Soc.*, **27,** 7, and private communication
KATZ, L., and PENFOLD, A. S., 1952, *Rev. Mod. Phys.*, **24,** 28.
MCMILLAN, E. M., 1945, *Phys. Rev.*, **68,** 143.
PIERCE, J. R., and SHEPHARD, W. G., 1947, *Bell Syst. Tech. J.*, **26,** 460.
REDHEAD, P. A., LECAINE, H., and HENDERSON, W. J., 1950, *Canad. J. Res.* A., **28,** 73.
VEKSLER, V., 1945, *J. Phys. U.S.S.R.*, **9,** 153.

Fundamental Processes of the Initiation of Electrical Discharges

BY C. GREY MORGAN AND D. HARCOMBE*
Department of Physics, University College of Swansea

Communicated by F. Llewellyn Jones; MS. received 7th April 1953

Abstract. The role of electrode and gas processes in the production of electrons which initiate spark discharges is investigated. The results for metal surfaces covered with thick ($\gtrsim 10^{-5}$ cm) oxide layers are consistent with the view that electrons may be liberated from sites of negative ions on the oxide. For metals covered with thin ($\sim 10^{-7}$ cm) tarnish films the emission is also field dependent but in these cases involves the action of a surface charge of positive ions in setting up an intense field across the film causing tunnel effect emission: in both cases the macroscopic applied electric field was of the order of 10^5 v cm^{-1}. The influence of surface micro-geometry on the cold cathode emission processes is examined. These experiments in dry air, hydrogen and argon ($p \sim 760$ mm Hg) indicated that initial electron production by detachment processes and the quenching of metastable atoms does not play a significant part in the discharge initiation. The influence of moisture on these processes is examined. Recent work on breakdown of liquid dielectrics is discussed in the light of the present work.

§ 1. INTRODUCTION

THIS paper describes a study of the fundamental processes involved in the initiation of discharges in dielectric media, and investigates the influence of the nature of the electrode surface on cathode emission as well as the influence of the nature of the gas on the action of ions in their role of supplying initiatory electrons. The method of using spark time lag data to obtain information about the mechanism of production of initiatory electrons in spark gaps in gases has recently been the subject of a rigorous theoretical and experimental analysis by Llewellyn Jones and de la Perrelle (1953), and the present investigation follows that basic work.

In order to produce a discharge in a spark gap an initiatory electron must be present in a suitable position when a sufficiently high potential difference is applied. In the absence of such an electron a voltage in excess of the static sparking potential V_s can be maintained for some time across the gap without the occurrence of a spark. This time is referred to as the total time lag of sparking. It consists of the initiatory or statistical lag and the formative lag. The initiatory lag is the time which elapses between the application of a potential equal to or greater than the static sparking potential and the appearance of the electron which leads to breakdown. The formative lag is the time taken for the development of the Townsend avalanche and the setting up of the secondary processes. Previous work with various gases has shown that the formative lag ($\sim 10^{-8}$ sec) is usually negligibly small compared with the mean statistical lag (~10^{-4} sec) for the short gaps (~0·03 cm) used in this work, and it is convenient to regard the statistical lag as being equal to the total lag. The following is the principle of the method of investigation.

---

* Now at Ministry of Supply, Sevenoaks.

§ 2. EXPERIMENTAL TECHNIQUE

The experimental technique involves the establishment of an electric field in a spark gap on a large number N of successive occasions, maintaining it constant each time for a period (~10^{-2} sec) which is long compared with the average time (~10^{-4} sec) taken for an electron to appear for many surfaces investigated. The time which elapses between the setting up of the field and the appearance of the electron, i.e. the initiatory lag, is recorded.

An electron once liberated has a certain probability W of leading to a spark, thus the time lag depends jointly upon the probability of electron liberation and upon W. In general W is a complicated function of position, electric field X_m, gas pressure p, and the Townsend ionization coefficients α and ω/α; in general W will be different for different electrons. However, in a parallel plate gap it will be the same for all electrons liberated at the cathode under a constant uniform electric field. This is a particular form of the general case considered by Wijsman (1949) and permits calculation of the numerical values of W as a function of gap voltage V from known values of the Townsend ionization coefficients. Figure 1 gives W as a function of V calculated for the particular cases of aluminium electrodes in air, argon and hydrogen; the electrode separation was 0·03 cm. It can be seen that for V greater than 1·35 V_s, W may be taken as unity in argon and hydrogen, and for V greater than 1·25 V_s in air. Under these conditions every electron released at the cathode leads to a spark; thus counting sparks is the same as counting the initiatory electrons.

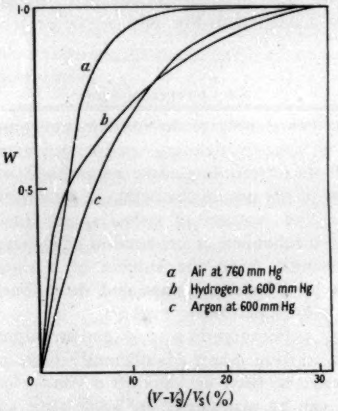


Fig. 1. Probability W as a function of percentage over-voltage. The curve for air at 600 mm Hg would approach those for argon and hydrogen.

If I the electron emission per unit time from the cathode is constant and $W=1$, the probability of a spark time lag of length t to $t+dt$ is $I \exp(-It)\,dt$ (Llewellyn Jones and de la Perrelle 1953); and the mean time lag $\bar{t}$ is given by

$$\bar{t}=1/I. \qquad \ldots\ldots(1)$$

It also follows that the standard deviation σ of these time lags is equal to the mean $\bar{t}$: thus determination of $\bar{t}$ from the measurement of N lags gives directly the rate of electron emission I. The probability $\exp(-It)$ of a lag exceeding t is determined experimentally by observing the number of times n in which the lag actually does exceed t out of a very large number N of trials.

Thus $$n/N=\exp(-It) \qquad \ldots\ldots(2)$$

hence $$I=(1/t)\ln(N/n). \qquad \ldots\ldots(3)$$

This relation provides a convenient method of obtaining I from a graph of $\ln(N/n)$ plotted against t. The linearity of such a graph together with the equality of σ and $\bar{t}$ all determined from experiment provides a rigorous check of the method.

§ 3. APPARATUS AND EXPERIMENTAL METHOD

A block diagram of the apparatus is given in fig. 2. The sequence of events for the measurements of a single spark lag was as follows. A trigger pulse from the source was applied simultaneously to the time base, calibrating oscillator and delay circuit; the time base and oscillator functioned immediately but the operation of the electronic switch was delayed for a few microseconds by the delay circuit. On closure, this switch connected the pulse generator to the gap for a period of 2 msec. This applied an impulse of potential difference of sensibly rectangular form and of sufficient magnitude to make $W=1$ to the gap for 2 msec, after which it was removed. The circuit then returned to its quiescent state to await the application of another trigger pulse at a pre-determined instant. The form of potential difference across the gap was displayed on a cathode-ray tube

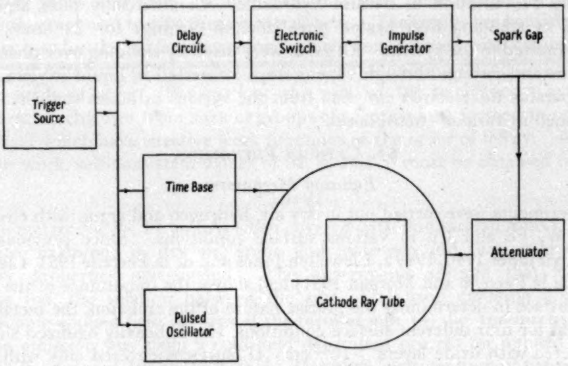


Fig. 2. Block diagram of apparatus.

screen after division in the attenuator. The vertical portion of the rectangular impulse was masked on the screen to leave visible only the horizontal trace. If no electron is liberated from the cathode during the application of the impulse the length of the trace is equal to the impulse duration of 2 msec. On the other hand, if an electron is liberated at a time t_1 after the electric field is established, breakdown immediately occurs (since $W=1$) and the potential difference immediately collapses, terminating the horizontal trace. The length of this trace in terms of the calibrated time base gives the statistical time lag t_1. This process is repeated after a pre-determined rest time by the application of a second trigger pulse and a second lag t_2, in general different from t_1, recorded. Photographic film was made to move vertically in a camera in front of the screen so that the lags were recorded separately as a series of horizontal straight lines. The records were analysed by various optical or photomultiplier arrangements. For quick assessment of the activity of a surface, direct visual observation was highly convenient. Owing to the ease and rapidity with which measurements could be made and analysed by this technique some millions of time lags were observed and recorded in the present investigation.

The variable gap discharge chamber, gas and vacuum system used were the same as those described and used by Llewellyn Jones and Grey Morgan (1953) but with the addition of a number of small spark gaps sealed in hard glass envelopes and connected to the system. The electrodes used in the variable gap discharge chamber consisted of cylinders 1·25 cm in diameter and 3 cm long of the materials to be examined. Hydrogen was generated electrolytically from a boiling 5% solution of barium hydroxide. Commercially pure argon was purified by standing over heated pure calcium turnings for 24 hours before being admitted to the system. All gases were dried by standing over phosphorus pentoxide and passage through vapour traps immersed in liquid oxygen. With this apparatus the electron emission from the various cathodes could readily be determined in various atmospheres.

§4. CATHODE PHENOMENA

Emission Mechanisms

Experiments were carried out in dry air, hydrogen and argon with electrodes of Ni, W, Fe and Cu in various surface conditions. Since previous work (Llewellyn Jones 1946, 1949 a, Llewellyn Jones and de la Perrelle 1951, Llewellyn Jones, de la Perrelle and Morgan 1951) had shown the importance of the nature of the surface in determining the precise nature of the emission, the metals were examined for four different surface conditions, viz. I, heavily oxidized surfaces, i.e. covered with oxide layers $\gtrsim 10^{-5}$ cm; II, surfaces covered only with a thin tarnish film $\sim 10^{-7}$ cm; III, smooth and clean surfaces; IV, electro-polished surfaces.

The electrodes in class I were those obtained by sparking and by long exposure to the atmosphere; such surfaces are of considerable interest in electrical engineering. Class II surfaces were highly polished first to grade 0000 emery paper and then with micro-alumina after which the surface was exposed to the atmosphere for the few minutes only required for the assembly of the discharge chamber which was then evacuated. Class III surfaces were obtained by the further conditioning and hydrogen ion bombardment of electrodes of class II. The fourth class of electrodes were obtained by electrolytic polishing.

Class I Surfaces

The measurements of the rate of electron emission I from cathodes of Ni, W and Fe which were covered with comparatively thick ($\gtrsim 10^{-5}$ cm) oxide layers showed that I measured in electrons per second was related to the applied field X_m by an equation of the form

$$I = AX_m^2 \exp(-D/X_m), \qquad \ldots\ldots(4)$$

where A and D are constants.

From the experimentally obtained values of A and D, the area S_1 and work function ϕ_1 of the source of electrons can be obtained by comparison with the Fowler–Nordheim (1928) equation for field emission. This theoretical equation may be written as

$$I = 38{\cdot}5 \times 10^{12} \frac{\zeta^{1/2}}{(\zeta+\phi_1)\phi_1^{1/2}} S_1 X^2 \exp\left(-6{\cdot}8 \times 10^7 \frac{\phi_1^{3/2}}{X}\right), \qquad \ldots\ldots(5)$$

where ζ is the Fermi energy of electrons and X is the microscopic field at the cathode surface and is a multiple of X_m given by

$$X = MX_m, \qquad \ldots\ldots(6)$$

where M is a factor representing the local intensification of the macroscopic fields by microscopic roughness. The values of ϕ_1 and S_1 thus obtained are $\phi_1 \sim 0{\cdot}5$ ev and $S_1 \sim 10^{-14}$ cm^2 where M is taken to be 10. These results are very similar to those obtained by Llewellyn Jones and de la Perrelle (1953) and by Llewellyn Jones and Grey Morgan (1953), which they completely confirm; there is thus no need to give them in detail here. The results are consistent with the view that for the surfaces examined which were covered with comparatively thick ($\gtrsim 10^{-5}$ cm) oxide layers, the initiatory electrons are produced under applied electric fields of about 10^5 v cm^{-1} at the rate of 10^3 to 10^5 sec^{-1} by a field dependent mechanism from sites or groups of negative ions formed in or on the oxide layer which have effective work functions of the order of 0·5 ev. As in the previous work, self-consistent values of M, ϕ_1 and S_1 must be obtained together.

Class II

The metals Ni, W, Fe and Al when covered with thin tarnish films (10^{-7} cm) were also examined in dry air, argon and hydrogen and their cold electron emissive properties investigated. The experiments showed that for these surfaces too, the relation between the rate of electron emission I and the macroscopic applied field X_m was of the form of eqn. (4), i.e. the mechanism of electron emission was again a cold field dependent process for applied electric fields of the order of 10^4 to 10^5 v cm^{-1}. However, consideration of the estimates of work functions and emitting areas obtained by comparison of the experimental data with the Fowler–Nordheim equation (5) showed that the detailed mechanism of emission could not in fact be pure field emission. The results obtained with tungsten electrodes, for example, lead to the following estimates for ϕ_1 and S_1: $\phi_1 = 0{\cdot}08$ ev, $S_1 = 3 \times 10^{-17}$ cm^2. Similar estimates were obtained for Ni, Al and Fe. A value of 3 was assigned to M in these cases, this value being not unreasonable for surfaces which have been highly polished and conditioned (Llewellyn Jones and Grey Morgan 1953). These low values of ϕ_1 and S_1 can clearly have no physical significance because the value of ϕ_1 is of the order of thermal energies which would give thermionic emission at room temperature, and the value of S_1 is less than atomic dimensions: the analysis, using (5), appears to be invalid and it is necessary therefore to examine more closely the conditions under which the electron emission took place.

It was noted in these experiments (which were carried out in dry gases) that considerable over-voltages could be maintained across the discharge gaps for times of the order of minutes before breakdown occurred. Once an initial breakdown occurred, however, all subsequent discharges could be readily initiated at much lower voltages and with time lags of the order of 10^{-4} sec or less. It follows therefore that ionization produced in the gap by the first measuring spark was effective in initiating sparks under subsequent impulses of e.m.f. Elementary considerations of the diffusion, drift, recombination and attachment of electrons and ions in the gas show that the ion concentration would be reduced to negligible amounts in the interval ($\sim 10^{-2}$ sec) between the application of successive impluses, and it may be concluded that the presence of post-breakdown ions *in the gas* did not play a significant part in the production of initiatory electrons. This view is strengthened when it is remembered that the experiments showed that the relationship between the rate of initiatory electron production I, the applied field X_m, and the cathode material was of the form which indicates that the mechanism was a cathode effect and was field dependent. It is not easy to see how any gas process could satisfy the experimental data. Yet the results obtained with dry gases indicate that post-breakdown ions were active in initiating discharges, because the presence of ions eliminated the very long (of the order of a minute) initial time lag. The mechanism of emission therefore has the properties of being: (*a*) cathode dependent, (*b*) field dependent, and (*c*) requiring the action of positive ions on the cathode.

A process which is consistent with these conditions has been proposed by Llewellyn Jones and Grey Morgan (1953) and is as follows. If the cathode has a thin insulating surface film, positive ions produced in the first spark can accumulate there, and set up an intense field X_1 across the film and liberate electrons from the underlying metal surface. The form of the potential energy barrier at the electrode surface is distorted by the presence of the field X_1 due to the ionic charge. The energy system is then the same as that considered by Stern, Gossling and Fowler (1929), and the rate of electron extraction is given by the relation

$$I = 38{\cdot}5 \times 10^{12} \frac{\zeta^{1/2} S_1 X^2}{(\zeta+\phi_1)\phi_1^{1/2}} \exp\left(-6{\cdot}8 \times 10^7 \frac{\phi_2^{3/2}}{X}\right) \times \frac{\phi_1}{\phi_2} \exp\left(-6{\cdot}8 \times 10^7 a \left\{\frac{\phi_1+\phi_2+(\phi_1\phi_2)^{1/2}}{\phi_1^{1/2}+\phi_2^{1/2}}\right\}\right), \qquad \ldots\ldots(7)$$

where $X = MX_m$ is in v cm^{-1}, and ϕ_1 and ϕ_2 are in ev and

$$\phi_1 - \phi_2 = aX_1, \qquad \ldots\ldots(8)$$

where a is the film thickness.

It is important to note that eqns. (7) and (4) have the same analytical form. Equation (7) is applicable to the present case because the form of the potential energy barrier at the electrode surface is identical when the field X_1 is set up by positive ions and when it is due to a contact potential difference $\phi_1 - \phi_2$ caused by the presence of a thin film of material of work function ϕ_2 on the base metal ϕ_1, as envisaged by Stern *et al.* All that is necessary to apply this equation to the present case it to remember that ϕ_2 now merely gives the surface field X_1 due to the ionic charge by the relation (8).

It can be clearly seen now that if the surface field due to positive ions is ignored then the estimates of the emitting areas given by eqn. (5) will be too small by a factor which is the inverse of the second exponential term in eqn. (7). If the film thickness were 10^{-7} cm in these experiments then the magnitude of this term is about 10^9, and the emitting areas calculated from eqn. (7) then become about 10^{-10} cm^2 or more. That such films can be present has been shown by Mott (1947), who assigned a value of 2×10^{-7} cm for the thickness of the oxide on aluminium, for example, after exposure to air.

Thus for the surfaces covered by thin ($\sim 10^{-7}$ cm) insulating tarnish films, the present experiments are consistent with the view that the initiatory electrons were liberated from the base metal under the influence of an intense electric field set up by the presence of positive ions on the insulating surface film. The maintenance of the surface positive charge will depend upon the properties of the thin tarnish film. In a number of experiments with different cathode surfaces, and therefore different tarnish films, the positive charge remained longer (as judged on this view by the enhancement of the field emission) on some films than on others, being particularly long for aluminium oxide. It might be expected that the decay of charge from the surface should depend on the resistivity of the tarnish film, and the density of the charge remaining on the surface should therefore decrease with the time allowed for decay. This is in accord with observations, because, when the time between successive pulses was increased from 10^{-2} sec to times of the order of minutes, the enhancement of the field emission decreased. It is of interest to note that with nickel and aluminium the decay was less than with the other metals tried, and enhanced emission was found even after delays of minutes. The presence of traces of moisture in the gas atmosphere greatly effected the rate of the decay, and this effect will be discussed in § 5 below; the observations are in accordance with the view that traces of water vapour in the gas reduced the insulation of the thin film.

Class III

When the surfaces of class II are subjected to prolonged conditioning and to ion bombardment in a hydrogen glow discharge the thin tenacious film is gradually removed (Llewellyn Jones and Davies 1951). Measurements made at successive stages during the removal of the film showed that for iron, for example, the emission decreased rapidly from about 10^5 electrons sec^{-1} at $X_m \sim 10^5$ v cm^{-1} until the rate of electron production corresponded to that due to natural processes only, i.e. there was negligible electron emission from the cathode. This shows clearly that the presence of the thin surface film is essential in producing the enhanced emission, presumably by providing a base upon which the positive ions can reside to set up an intense field at the substrate metal surface. Removal of the film destroys this high surface field leaving the macroscopic applied field which is then too low to extract electrons from the cathode, in which case the electrons which initiate discharges are those produced by natural processes.

While the electron emission from iron and copper was considerably reduced as a result of the bombardment by hydrogen ions in a glow discharge, no significant reduction was obtained with aluminium electrodes, although the treatment was continued for over one hundred hours. Whereas the surfaces of iron and copper were obviously cleaned and made brighter by this treatment, the aluminium cathode developed a dull black layer, i.e. the thin surface oxide film was not removed; instead a much thicker contaminating layer, possibly a hydride, was formed. The glow discharge between the Al electrodes was not uniform as was the case for copper and iron: there was a tendency for the formation of high current density discharges which stood out against the background of a very faint glow. These seemed to roughen the surface. The combined effect of the formation of the black contaminating layer and the roughening of the cathode gave rise to conditions which appeared to give enhanced electron emission of the order of 10^4 electron sec^{-1} at low electric fields of about 10^4 v cm^{-1}. Similar behaviour of aluminium in hydrogen glow discharges is reported by Fan (1939).

Surface Micro-Geometry

It is sometimes considered that electro-polished surfaces are mainly atomically smooth; if this were so there would be no intensification of the local field due to microscopic irregularities for such surfaces, and the value of M should be unity.

In previous work (Llewellyn Jones and Grey Morgan 1953) it has been shown that the emission from an electrolytically polished copper surface did not differ greatly from that from a standard surface which had been carefully hand polished and subjected to a conditioning process by light current sparking.

To compare further the properties of electro-polished and hand-polished surfaces, aluminium electrodes were electrolytically polished in accordance with the method described by Evans and Whitram (1947). These electrodes were mounted in the discharge chamber and dry hydrogen was admitted to a pressure of 600 mm Hg. The emission obtained from these electrodes was compared with that from aluminium electrodes which had been hand-polished and conditioned. The results of this comparison are given in the table from which it is seen that the emission from the electrolytically polished surfaces was found to be greater by a small amount than from the conditioned surface under the same electric fields. Thus the two types of treatment resulted in surfaces which had approximately the same degrees of smoothness.

Electric field (v cm$^{-1}\times10^4$)		4·19	4·94	5·43
Emission (electron sec$^{-1}\times10^{-3}$)	electro-polished	7·64	17·4	23·4
	hand-polished	9·0	15·0	19·5

Further careful hand polishing and prolonged conditioning of aluminium produced a much more inactive surface than was obtained by electrolytic polishing. This result is illustrated in fig. 3 in which the slope of the ($\ln N/n, t$) graphs gives the emission I. It can be seen that for the same value of electric field, the emission from the conditioned surface was only about one tenth of that from the electro-polished surface. This result was again confirmed in experiments with nickel which had been electrolytically polished in a 70% solution of sulphuric acid at a current density of 0·25 A cm^{-2}. In this case the emission from the electropolished Ni exceeded that from the conditioned Ni surfaces by an even greater quantity. Further experiments with electro-polished surfaces established that the emission was, in fact, field emission and related to the applied field by eqn. (4). This is shown by the linear relation obtained between $\ln I/X_m{}^2$ and $1/X_m$ in fig. 4 which was derived from the data of these experiments. These results are consistent with the view that electro-polishing of surfaces does not necessarily always produce a high degree of surface smoothness. On the contrary, it is likely that cliff-like crystal edges are revealed (Evans and Hopkins 1952). These edges can lead to intensification of the local electric field and give rise to

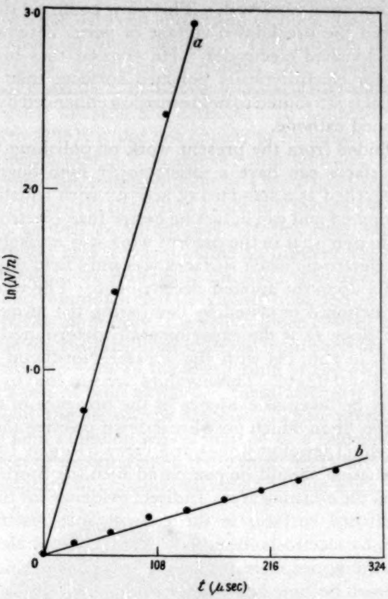


Fig. 3. Electron emission from (*a*) electro-polished and from (*b*) conditioned aluminium electrodes.

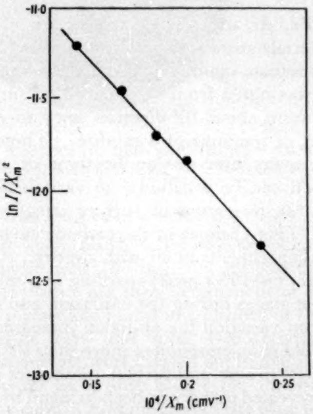


Fig. 4. Dependence of electron emission on electric fields for electro-polished aluminium.

the enhanced field emission observed. This view confirms the work of Hadden (1951) who measured the breakdown voltage *in vacuo* between electro-polished and between hand-polished electrodes. He showed that breakdown occurred at a lower voltage for electrolytically polished surfaces than for hand-polished surfaces. This effect is attributed to field emission enhanced by surface roughness on the electro-polished cathode.

It may be concluded from the present work on polishing that hand-polished and conditioned surfaces can have a satisfactorily high degree of smoothness, and they may be regarded as a standard of activity with which other more active surfaces can be compared and can in fact be better than electro-polished surfaces.

It is important to note that in the present work it is unlikely that the emission observed from the electro-polished surfaces was pure field emission in the sense that it depended only upon the applied electric field. The values of the emitting area S and work function ϕ obtained by comparing the numerical values of the constant A and the slope D of the experimentally determined graph of $\ln I/X_m{}^2$ against $1/X_m$ (fig. 4) in eqn. (4) with the Fowler–Nordheim equation (5) were $S \sim 10^{-19}\,\text{cm}^2$ and $\phi \sim 0{\cdot}02\,\text{ev}$. These values are too low to have any physical significance and may be taken as evidence of the presence of a thin ($\sim 10^{-7}$ cm) surface insulating film upon which post-breakdown positive ions accumulate and set up an intense field (Llewellyn Jones and Grey Morgan 1953). In this case the experimental equation should be compared with the Stern–Gossling–Fowler equation (7) to obtain the emitting area. Indirect evidence for the presence of such films on electro-polished surfaces is the considerable resistance to corrosion shown, for example, by electro-polished iron. Samples of electro-polished iron remain rust free in the atmosphere for several weeks while unpolished samples from the same specimen become heavily rusted in a few days. The presence of a thin protective film can probably account for this important fact (Evans and Hopkins 1952).

Electrode Growths

Experiments with Fe, Al and Cu electrodes are of considerable interest because with these materials surface instability appeared to occur; the cathode emission was found to fluctuate rapidly. Again, considerably enhanced emission of a spasmodic nature was noted from W, Fe and Al surfaces. In these cases the emission increased from about 10^4 electron sec^{-1} to greater than 10^6 to 10^7 for periods of the order of fractions of a minute. These effects can be of the greatest importance in many practical applications of spark ignition devices. This spasmodic emission can be attributed to variations in the nature of the electrode surfaces and the formation of surface irregularities which intensify the local electric field. That changes in the cathode surface can take place was clearly demonstrated in experiments in air with copper. Copper electrodes were oxidized by heavy current ($\sim$ 100 A peak) sparking in a subsidiary apparatus and the emission measured at stages during the oxidation and erosion of the surface. After some initial random variation the emission increased rapidly and voltages 5% in excess of V_s resulted in an emission of more than 10^6 electron sec^{-1}. Quite suddenly this emission fell to zero and further discharges could not be initiated. Examination of the gap revealed that it had been bridged by a growth (fig. 5) which was practically the segment of a cone in shape. This can be compared with growths obtained by Compton, Mendizza and Arnold (1951). Measurements showed that the resistance between anode and cathode was a fraction of an ohm as the growth had short circuited the gap. This growth, and others obtained subsequently on copper, appeared so quickly that visual observation of its development was not possible. However, it seemed that the growths developed on the anode and grew towards the cathode. In contrast to similar growth observed by Haworth (1951) it did not appear to be molten.

In other experiments designed to remove thin tenacious surface films using the method of hydrogen ion bombardment in a glow discharge of 10 mA current density at 10 mm Hg pressure (Llewellyn Jones and Davies 1951), similar growths were found on iron as well as on copper. In this case, however, the rate of growth was much slower and it was possible to observe the development from the anode to the cathode. On iron the shape was not as clearly defined as for the growths on copper and resembled that previously obtained on tungsten electrodes under 100 A sparks in this laboratory (Llewellyn Jones 1949 b) and had a resistance of a few ohms. The micro-transfer of material from such anode growths, which are friable, to the cathode leads to the sharpening of the projection and cathode irregularities. This can be seen in fig. 5. Small irregular pieces of the growth have become detached and lie on the cathode. Their presence results in considerably enhanced field emission which may be of a spasmodic nature owing to the conditioning effect of sparking upon these small irregularities. This accounts for some of the instability of emission observed in this work.

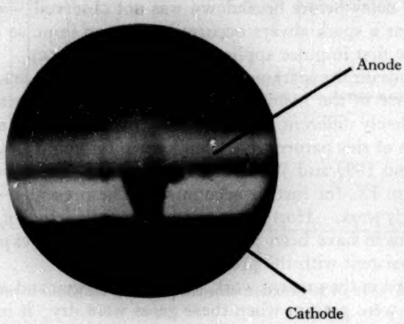


Fig. 5. Growth bridging a copper electrode gap (the dark parts on the cathode are reflections of the bridging growth and anode). Electrode separation = 0·03 cm.

§ 5. GAS PROCESSES INVOLVED IN INITIATION

When conditions are such that initiatory electrons cannot be liberated from the cathode, the discharge must be initiated by electrons released in the gas. Possible initiatory mechanisms are then detachment of electrons from negative ions, the quenching of metastable atoms in collisions of the second kind and also ionization due to cosmic rays and local radioactivity. To investigate the processes involved, experiments were carried out in dry and moist gases using the same general experimental technique as for the investigation of the cathode processes; and also by a careful examination of the time lags in the measurement of the static sparking potential V_s. For this part of the investigation the static potential across the gap was increased in steps of ten volts, each step being maintained for one minute; if no spark occurred in this period the potential was increased by a further ten volts, and so on until a spark occurred.

The gases were thoroughly dried by standing over phosphorus pentoxide for at least 24 hours and by passage through vapour traps immersed in liquid oxygen. The results obtained may be summarized briefly as follows:

(*a*) In the dry gases, measurements of V_s under static conditions showed that potentials of up to $4V_s$ could be sometimes maintained for periods of minutes before the occurrence of breakdown, but no such potentials could be maintained immediately after the first spark.

(*b*) In the dry gases, although, in the application of a succession of impulses of 10^{-2} sec duration, potentials of up to $4V_s$, and in one case $5V_s$, were maintained across the gap during periods of minutes before breakdown occurred; once this initial breakdown had occurred subsequent time lags were all quite short even though the potential was reduced to a little above V_s: the emission corresponding to these lags follows the field law discussed above (eqn. (4)).

(*c*) The introduction of moisture to the dry gas immediately destroyed these conditions. The long initial delay was at once ended, and breakdown occurred immediately on the application of an impulse. Breakdown was also produced under the successive e.m.f. impulses but the lags were longer than those found in the dry gases.

(*d*) In measurements of V_s under static conditions in a moist gas a long ($>10^{-2}$ sec) initial delay before breakdown was not observed.

(*e*) In moist gas a spark always occurred under an impulse of e.m.f. within the duration of the first impulse applied (i.e. within 10^{-2} sec).

The observations on the sparking potential measurements are illustrated in an approximate manner in the histograms of fig. 6. The scales are so drawn to indicate the completely different range of times involved in dry and moist gases.

General results of this nature have been known for about a century (Faraday 1838, J. J. Thomson 1893 and Warburg 1897) (see also Thomson and Thomson 1933, Vol. II, chap. IX, for further accounts). The present results confirm in every detail this early work. However, no complete explanation of these important phenomena appears to have been published to date, but an explanation can be given which is consistent with the present observations.

In the gases used in the present work, viz. air, hydrogen and argon, it is likely that free electrons were present when these gases were dry; it is thus necessary to account for the long delay which occurs before breakdown both in impulse and static conditions even though free electrons are present in the gap. The probable explanation is as follows: the transit time of electrons across the short gaps used (0·03 cm) is small ($<10^{-8}$ sec) compared with the time taken to establish the final value of the field in the gap, and free electrons would be swept to the anode when the parameter X_m/p was still low, and thus no ionization would be produced by the motion of these electrons; i.e. the gap would be cleared of electrons long before the voltage passed the breakdown value. In this case it is therefore necessary to await the fortuitous liberation of an electron near the cathode by natural processes *after* the field has reached sparking values in order to initiate a spark. In gaps of the size used this would involve a delay which could be as long as a few minutes, i.e. of the order of magnitude observed.

In a moist gas, on the other hand, owing to the formation of large numbers of very low mobility negative ions by the attachment of electrons to water molecules, ions are likely to remain in the gap for much longer times and be present when the voltage attains and even passes the minimum breakdown value. If electron detachment then occurs, breakdown would follow immediately. In this work X_m/p always exceeded 85 v cm^{-1} mm^{-1} Hg after the final field was established and the absence of a long delay ($\sim 10^{-2}$ sec) in a moist gas before breakdown is accounted for if electron detachment occurs within this interval.

It should also be noted that in gaps which have been exposed to a moist atmosphere for some time there is a high probability of cold cathode emission from sites of negative ions at the cathode surface as discussed above (Llewellyn Jones and de la Perrelle 1951, 1953, Llewellyn Jones, de la Perrelle and Morgan 1951). In long gaps, however, where the sparking field may be too low to produce this enhanced field emission, initiation by electron detachment from a water molecule is the probable mechanism. In either case there would be no long delay between the application of the potential difference and the occurrence of a spark, and this is in agreement with the observations. A further conclusion follows from this work: in hydrogen there are no metastable states, thus discharge initiation by electrons liberated by the quenching of metastable atoms is not possible in this gas; however, such a process is permitted in argon which has metastable levels. Since there was no marked difference between the rates of discharge initiation in these two gases it follows that the production of electrons from collisions of the second kind between gas molecules and metastable atoms does play a significant part in discharge initiation. Over the range of pressures 200–760 mm Hg there appeared to be no significant dependence of emission upon pressure; the nature of the gas was the important factor influencing the emission mechanism.

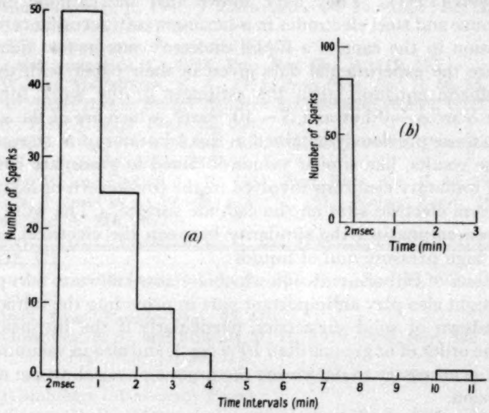


Fig. 6. Distribution of time lags obtained in measurements of V_s in (*a*) dry gas, (*b*) moist gas.

Discharge Initiation in Liquid and Solid Dielectrics

So far in this paper there has been considered only the initiation of discharges in gases, but there is no reason why the processes of field emission envisaged should not be operative in breakdown in liquid or even solid dielectrics.

This cold extraction of electrons at remarkably low electric fields has important applications to the electrical breakdown of gases at very high pressures, when high ($\sim 10^5$ to 10^6 v cm^{-1}) electric fields obtain, since the field emission can lead to a lowering of the sparking potential through the failure of Paschen's law (Llewellyn Jones and Grey Morgan 1951).

At the very highest pressure these processes of cathode emission are likely to be the predominant ionization processes (Trump, Cloud, Mann and Hanson 1950) and it seems probable that this might be operative in the breakdown of liquid dielectrics. There would appear to be no reason on general grounds why the mechanism of electron emission when a metal electrode is in contact with a gas should not be the same as when the metal is in contact with a liquid dielectric, or even a solid dielectric. It is interesting therefore to discuss measurements of cathode emission for the case of ionization currents which occur prior to the breakdown of a liquid dielectric. It is possible to do this because of the recent comprehensive investigation of conduction and breakdown in liquids by Goodwin and Macfadyen (1953). They have shown that the cathode emission from phosphor bronze and steel electrodes in *n*-hexane was in accordance with eqn. (4): i.e. the emission in the case of a liquid dielectric was in fact field dependent. If we compare the experimental data given in their paper with the theoretical Fowler–Nordheim equation, then the estimates of the work function ϕ and emitting area S are $\phi \sim 0{\cdot}1$ ev and $S \sim 10^{-15}$ cm^2, which are of the same order of magnitude as those previously obtained in this laboratory for gaseous dielectrics.

Now these results, like similar values obtained in gases, are consistent with the view that initiatory electrons involved in the breakdown of liquid dielectrics are emitted from electron sites on the cathode surface. The work of Goodwin and Macfadyen emphasizes the similarity between the electrical breakdown of gases at very high pressure and of liquids.

The processes of cathode emission which are now known to take place in gases and liquids might also play an important part in providing the initiatory electron for the breakdown of solid dielectrics, particularly if the intrinsic breakdown fields are of the order of or greater than 10^5 v cm^{-1}, and also in vacuum breakdown. Further work is necessary to determine the precise physical nature of the sites of electron emission.

§ 6. Electron Emission from Semiconductors

A preliminary study of cold emission from semiconductors has been made. The semiconductors were prepared for the authors by the Plessey Company Ltd., and consisted of (*a*) pure zinc oxide, (*b*) pure zinc oxide plus 0·1% by weight of copper oxide, (*c*) zinc oxide plus 10% by weight of nickel carbonate, in the form of cylinders 1·25 cm in diameter and 3 cm long with rounded edges at the flat surface. These specimens were made the cathode of a spark gap the anode of which was highly polished and conditioned nickel. The experimental procedure was exactly the same as that for metal cathodes.

It was found that the electron emission was considerably greater than that obtained from metals under identical conditions of electric field, gas pressure etc. For example with fields of the order of 10^4 v cm^{-1} and pressures approximately atmospheric, the emission exceeded 10^7 electron sec^{-1}. Further work with single crystals of germanium is now in progress.

ACKNOWLEDGMENTS

The authors wish to offer their best thanks to Professor F. Llewellyn Jones, under whose direction this work was carried out, for his most helpful advice and encouragement. They also wish to thank the Research Department of the Plessey Company Ltd., Caswell, Northants, for preparing electrodes of semiconductors in the form required in this work.

They wish to express their gratitude to the Scholarships Committee of the Institution of Electrical Engineers for the award to one of the authors (C.G.M.) of the I.M.E.A. Research Scholarship, which enabled him to take up this work, and to the Department of Scientific and Industrial Research for the award of a maintenance grant to the other (D.H.) during this work.

REFERENCES

COMPTON, K. G., MENDIZZA, A., and ARNOLD, S. M., 1951, *Corrosion*, **7,** 327.

EVANS, D. J., and HOPKINS, M. R., 1952, *J. Electrodepositors Tech. Soc.*, **28,** Paper 8.

EVANS, U. R., and WHITRAM, J., 1947, *J. Electrodepositors Tech. Soc.*, **22,** 22.

FAN, H. Y., 1939, *Phys. Rev.*, **55,** 769.

FARADAY, M., 1838, *Experimental Researches*, § 1417.

FOWLER, R. H., and NORDHEIM, L., 1928, *Proc. Roy. Soc.* A, **119,** 173.

GOODWIN, D. W., and MACFADYEN, K. A., 1953, *Proc. Phys. Soc.* B, **66,** 85.

HADDEN, R. J. B., 1951, *Atomic Energy Research Establishment Report*, G/M 92.

HAWORTH, F. E., 1951, *J. Appl. Phys.*, **22,** 606.

LLEWELLYN JONES, F., 1946, *Nature, Lond.*, **157,** 371, 480; 1949 a, *Proc. Phys. Soc.* B, **62,** 366; 1949 b, *Brit. Sci. News*, **2,** 263.

LLEWELLYN JONES, F., and DAVIES, D. E., 1951, *Proc. Phys. Soc.* B, **64,** 397.

LLEWELLYN JONES, F., and DE LA PERRELLE, E. T., 1951, *Nature, Lond.*, **168,** 160; 1953, *Proc. Roy. Soc.* A, **216,** 267.

LLEWELLYN JONES, F., DE LA PERRELLE, E. T., and MORGAN, C. G., 1951, *C.R. Acad. Sci., Paris*, **232,** 572.

LLEWELLYN JONES, F., and MORGAN, C. G(REY), 1951, *Phys. Rev.*, **82,** 970; 1953, *Proc. Roy. Soc.* A, **218,** 88.

MOTT, N. F., 1947, *Trans. Faraday Soc.*, **43,** 429.

STERN, T. E., GOSSLING, B. S., and FOWLER, R. H., 1929, *Proc. Roy. Soc.* A, **124,** 699.

THOMSON, J. J., 1893, *Phil. Mag.* [5] **36,** 313.

THOMSON, J. J., and THOMSON, G. P., 1933, *The Conduction of Electricity through Gases*, Vol. II (Cambridge : University Press).

TRUMP, J. G., CLOUD, R. W., MANN, J. G., and HANSON, E. P., 1950, *Trans. Amer. Inst. Elect. Engrs.*, **69,** 961.

WARBURG, E., 1897, *Wied. Ann.*, **62,** 385.

WIJSMAN, R., 1949, *Phys. Rev.*, **75,** 833.

The Electrical Conductivity and Current Noise of Carbon Resistors

By I. M. TEMPLETON and D. K. C. MacDONALD*
Clarendon Laboratory, Oxford

MS. received 30*th March* 1953

Abstract. The variation of resistance with temperature, pressure and magnetic field in several types of carbon resistor has been measured down to helium temperatures. Temperature variation is interpreted on the basis of an assembly of contacts with two different ranges of barrier height. A resistance decrease with pressure of approximately 0·1% per atmosphere is found in carbon compound resistors, and a small size effect variation with magnetic field in the case of carbon film resistors. Noise measurements have been made which show that the $1/f$ law holds at all temperatures. The temperature dependence of the noise is not in agreement with Macfarlane's theory in terms of a single activation energy. Measured noise power is found to be closely proportional to I^2, while the ambient temperature noise level at 10^{-5} A for small resistors is expressible as $\overline{\delta R^2}/R^2 \simeq 10^{-11}\, \delta f/f$.

General Introduction

The work to be described was undertaken as part of a programme of noise measurement in semiconductors. It may conveniently be divided into two sections dealing first with resistance measurements, and secondly with the work on noise.

I. RESISTANCE MEASUREMENTS

§ 1. Introduction

Previous work on amorphous and graphite carbon has unfortunately been of a rather empirical nature, investigating its suitability for low temperature resistance thermometry, and little is apparently known of the detailed structure of carbon resistors or of the fundamental reasons for their physical behaviour. The only theories applicable to the temperature variation of resistance are perhaps those advanced by Mrozowski (1950, 1952) in connection with graphite of various forms, and no detailed agreement has been found with any of these.

§ 2. Experiments and Results

2.1. *Investigation of Contacts*

It was desirable first to check that the contact resistance of the external connections to carbon resistors is low, since high-resistance contacts might cause serious difficulty in obtaining accurate measurements of resistance and noise. Several one-watt resistors were therefore mounted in a jig which allowed the direct application of axial potential contacts, and measurements were made on a potentiometer at about 290°K, 90°K and 20°K, the lower temperatures being obtained by immersion in liquid oxygen and liquid hydrogen respectively. These measurements showed that any contact resistance was less than 0·1% of the whole and was therefore negligible.

---

* Now with the Division of Physics, National Research Council, Ottawa, Canada.

2.2. *Variation of Resistance with Pressure*

Since measurements in the helium range would involve changes of pressure, and the approach to helium temperatures in a modified Simon liquefier (Simon 1936) might involve pressures of the order of 100 atmospheres, a check was made that there was no very large resistance–pressure effect. Measurements were made with a Wheatstone bridge on specimens contained in the liquefier to be used. A carbon compound resistor showed a reduction of the order of 0·1% per atmosphere, as might be expected from its granular construction. No measurable effect was found in a carbon film resistor. The maximum error from this cause is thus 10%, which is negligible when compared with the effects due to temperature changes in the same range.

2.3. *Variation of Resistance with Temperature*

Measurements down to hydrogen temperatures made on Erie $\frac{1}{4}$-watt carbon compound resistors and plotted as R_T/R_{290} on a logarithmic scale against $1/T$ with the intention of investigating activation energies showed immediately two distinct groups of resistor (fig. 1). Information obtained from the manufacturers led us to believe that the lower values of resistor have some graphitic inclusions—possibly produced by a heating process during production. More detailed measurements down to helium temperatures on a 'graphitic' carbon compound

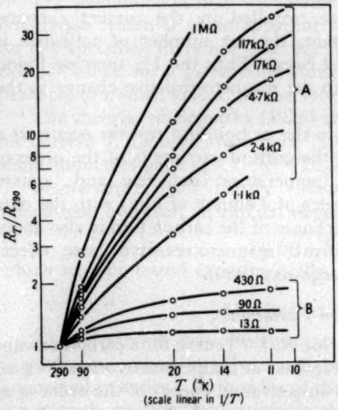


Fig. 1. Variation of resistance with temperature of a range of carbon compound resistors (nominal values quoted on curves). A, 'amorphous carbon' type; B, 'graphitic carbon' type.

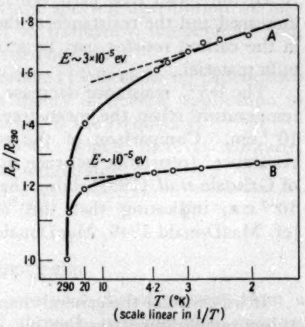


Fig. 2. Temperature variation of relative resistance of carbon resistors for estimation of activation energies. A, 100 kΩ (nominal) 'cracked carbon' (film) resistor; B, 10Ω (nominal) 'carbon compound' resistor.

resistor (10 Ω nominal value) and on a 100 kΩ carbon film resistor showed (fig. 2) apparently two activation energies, of the order of 10^{-3} ev or higher at temperatures between 300°K and 20°K and $\sim 10^{-5}$ ev at lower temperatures. It may be noted here that the carbon film resistor is formed of a pyrolitic deposit of carbon, and such films have been shown by X-ray analysis to contain graphitic groups together with 'gaseous' carbon molecules.

Comparison of these measurements with those made recently by Wilks and Webb of this laboratory in connection with carbon resistance thermometers has shown a striking resemblance between the curves for *all* types of carbon resistor, including those made in the laboratory from graphite dispersions. That is, all show more or less clearly two different activation energies, though there is a considerable range of values for these.

2.4. *Variation of Resistance with Magnetic Field*

A magnetic field of 8 kilogauss was applied to several specimens at various temperatures. A resistance rise of 1% was observed in the case of a 10Ω carbon compound resistor at 4·2°K, whilst a 100 kΩ carbon film resistor showed a drop of 0·5% at the same temperature. On the other hand, a specimen of polycrystalline graphite showed under the same conditions a magneto-resistive increase of some 146%.

§ 3. DISCUSSION

3.1. *Magnetic Field Experiments*

The results of the experiments with magnetic fields suggest that we may take as our model for a carbon compound resistor a number of low resistance granules with high resistance contacts between them. A magnetic field has no effect on a potential barrier, but can increase the resistance of bulk material by effectively increasing the average distance travelled by the current carriers. Since the mean free path remains constant, the net number of collisions is increased and the resistance of the material rises. Thus the 1% increase found in the carbon resistor may be attributed to the magneto-resistive change in the bulk material.

The 0·5% resistance decrease noticed in the carbon film resistor occurs at a temperature when the mean free path of the current carriers is of the order of 10^{-7} cm. Comparison of the resistor's temperature coefficient and 'square resistance' (resistance between opposite sides of a square of film) with the data of Grisdale *et al.* (1951) shows that the thickness of the carbon film is also about 10^{-7} cm, indicating that this is the normal magneto-resistive 'size effect' (cf. MacDonald 1949, MacDonald and Sarginson 1950).

3.2. *Temperature Variation*

If we consider the general shape of the (log R, $1/T$) curve for a carbon resistor at low temperatures, we find two distinct regions: at temperatures *below* 20°K an approximate straight line corresponding to an activation energy of the order of as low as $\sim 10^{-5}$ up to $\sim 10^{-3}$ ev, and *above* 20°K a portion which tends towards but does not reach a straight line corresponding to a higher activation energy. There is also at high temperatures a saturation region in which thermal scattering of conduction electrons in the bulk material causes the resistance to rise again, but we are not concerned here with this effect.

Assuming then that the resistance is located effectively at the contacts, we must envisage some contact mechanism whereby there are *two* different activation energies with a fairly wide range of possible values. It must be emphasized that the effects found can only be explained by a *parallel* mechanism, e.g. of two potential barriers in parallel. In a series arrangement, only that producing the *higher* resistance would be apparent in external measurements.

One possible explanation is that while most contacts involve barriers of height 10^{-3} to 10^{-2} ev, there is a small number of 'good' contacts of much lower barrier height. The height of these lower barriers, which in any one specimen appear to be of the same order, is presumably determined by the forming processes used during manufacture. Since graphite under certain conditions behaves as a semiconductor of zero activation energy (Wallace 1947), the lower activation energy could be interpreted as a parallel 'short circuiting' of contacts by graphitic inclusions.

II. NOISE MEASUREMENTS

§ 1. Introduction

The first detailed investigations of current noise since its discovery by Johnson (1925) were made by Christensen and Pearson (1936) using carbon microphones and other granular (e.g. carbon compound) resistances. This emphasis on the association of contacts with this form of noise led to the adoption of the name 'contact noise', but since the effect is observed also in bulk semiconductors such as germanium, the name is, perhaps, inappropriate. Other workers in this field, notably Abson *et al.* (1946) and Campbell and Chipman (1949) have confirmed over a limited temperature range and at various frequencies the general findings of earlier workers.

The phenomenon may with some simplification be regarded as a resistance fluctuation which is made apparent to us by passing a steady current; that is, for any specimen there is a quantity $\delta R^2/R^2$ (whose spectral distribution we may call ρ_ω) which for a given bandwidth depends on frequency, temperature and, perhaps, to a small extent, on the value of the current itself.

The theories of Schottky (1926) and Macfarlane (1950) show no dependence on current I, though Macfarlane (1947) in a theory of general application to semiconductors gave $\rho_\omega \propto I^{x-1} f^{-x} (1 \leqslant x \leqslant 2)$ where f is the frequency.

Macfarlane's later theory (Macfarlane 1950), which involves contacts, introduces, however, an explicit dependence on temperature:

$$\rho_\omega \propto T^{-2} e^{-E/8\boldsymbol{k}T} f^{-x} (x \simeq 1)$$

which may be tested experimentally.

§ 2. Experimental Procedure

The apparatus was designed so that a measured direct current from a carefully smoothed supply could be passed through two identical resistors connected in series. Both sides of the current supply were effectively earthed to a.c. signals so that any resistance fluctuations in the specimens produced an alternating potential relative to earth at the junction between the two resistors. This arrangement has two specific advantages apart from those normally inherent in a balanced system: (i) the available noise power is double that obtainable from a single carbon resistor in series with a wire-wound resistor of equal value;

(ii) each resistor, considered as a source of noise power, 'sees' in the other at all temperatures optimal a.c. loading.

The two resistors were contained in a small helium-filled copper capsule which could be surrounded with liquid oxygen or hydrogen. For measurements at helium temperatures the two resistors were mounted in a Simon (1936) liquefier as used in the resistance measurements. Connections to the specimens were made with fine insulated constantan wires passed down the german silver tube which supported the copper capsule.

The noise signal was fed via a low-capacity coaxial cable and a condenser of high insulation resistance into a low-noise preamplifier. A selector switch at the preamplifier input allowed readings to be taken of the thermal or 'Johnson' noise in the 1MΩ grid resistor which was then used as a reference level. From the preamplifier the signal was fed to a wave analyser of a nominal 4 c/s bandwidth. The output from this (a 50 kc/s signal modulated by the noise) was taken to an oscilloscope and from this to the detector, using a cathode-ray tube as a visual monitor. The detector was in some cases an envelope-detector diode and in others a vacuum thermocouple, both systems having a time constant of about 10 seconds. The detailed frequency response of the complete apparatus was determined by injecting a signal across a small resistor in the 'earthy' side of the bridge.

Using resistors of the carbon compound and carbon film types with room temperature values between 100 kΩ and 1 MΩ, measurements have been made of the variation of current noise with frequency, temperature and current. The effect of a magnetic field has also been examined.

§ 3. RESULTS

3.1. *Frequency Dependence*

In the range 20 c/s to 10 kc/s all specimens obeyed more or less exactly a $1/f$ law at currents and temperatures ranging from 1 μA to 100 μA and 290°K to 4·2°K respectively.

A typical spectrum is shown in fig. 3. It may be mentioned here that measurements at room temperature have recently been extended to frequencies lower than 10^{-3} c/s where the $1/f$ law is *still* obeyed rather precisely (Rollin and Templeton 1953).

3.2. *Current Dependence*

Up to 20 μA, at frequencies and temperatures as above, the measured mean square noise *voltage* in all specimens follows more or less exactly a law $\overline{\delta V^2} \propto I^{2\cdot 0}$ (though a reduction of the index of I to 1·9 could be read into some of the earlier results).

The same law continues to apply for a 100 kΩ carbon film resistor for currents as high as 500 μA. However, carbon compound types show, at currents ranging from 20 to 60 μA, a fall-off from a straight line on a graph of log $(\overline{\delta V^2})^{1/2}$ plotted against log I. These deviations are found to be non-reproducible (fig. 4 (both types)). It should be noted that the power dissipation involved is still far below the rated value of the resistor.

3.3. *Temperature Dependence*

ρ_ω was deduced for a carbon compound resistor and a carbon film resistor at temperatures down to 14°K and 1·5°K respectively. In each case the value at 1 kc/s (1 c/s bandwidth) with 10 μA flowing was taken.

Graphs were then drawn of log $T\sqrt{\rho_\omega}$ plotted against $1/T$ for comparison with Macfarlane's theory (fig. 5), which would predict a linear relationship.

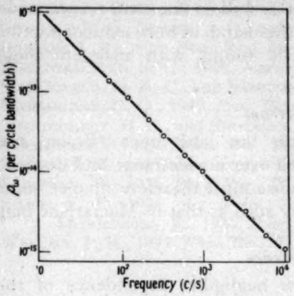


Fig. 3. Spectral distribution of current noise in 100 kΩ (nominal) carbon film resistor, $I=10^{-5}$ A, $T=290$°K. (Continuous line shows exact $1/f$ law.)

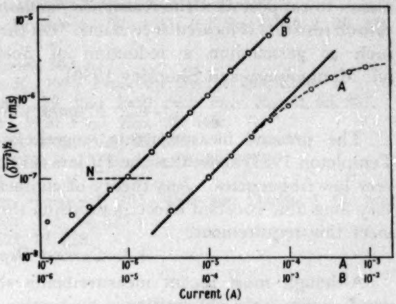


Fig. 4. Noise voltage (r.m.s.) from carbon resistors per cycle bandwidth ($T=290$°K). A, 120 kΩ (nominal) carbon compound resistor; B, 100 kΩ (nominal) carbon film; N, background noise level. Full lines give equation $(\overline{\delta V^2})^{1/2} \propto I^{1\cdot 0}$.

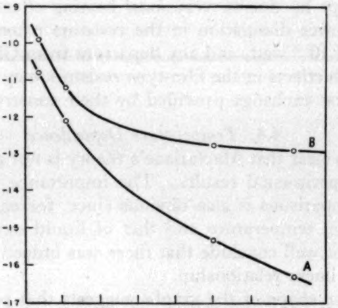


Fig. 5. Temperature dependence of current noise in carbon resistors for comparison with Macfarlane's theory. A, 120 kΩ (nominal) carbon compound resistor; B, 100 kΩ (nominal) carbon film resistor.

3.4. *Effect of a Magnetic Field*

A field of 8 kilogauss produced no measurable change of noise level in any specimen.

3.5. *General*

The value of ρ_ω for a current of 10 μA has been found to lie in the range 3×10^{-11} to 3×10^{-12} $\delta f/f$ for $\frac{1}{4}$-watt carbon compound and $\frac{1}{2}$-watt carbon film resistors.

§ 4. DISCUSSION

4.1. *Magnetic Field*

The fact that no measurable effect is produced by a magnetic field gives reason to suppose that the resistance *fluctuation* as well as the main resistance in carbon resistors is located at contacts. On the other hand, in bulk semiconductors such as germanium a reduction of noise is found with magnetic field (cf. Montgomery and Shockley 1950).

4.2. *Frequency Spectrum*

The present measurements, together with the later ones (Rollin and Templeton 1953) show that the $1/f$ law is obeyed over a wide range and down to very low frequencies. Any theory of contact noise must therefore involve some very long time constant effect: a diffusion theory such as that of Macfarlane may meet this requirement.

4.3. *Current Dependence*

Although most recent measurements show negligible dependence of the mean square (relative) resistance fluctuation on current, earlier work suggested a small dependence $\overline{\delta R^2}/R^2 \propto I^{-0\cdot 1}$. Such a reduction might possibly be associated with a progressive reduction in the number of fluctuating carrier-emitting centres; alternatively it might be attributed to the sweeping away of carriers from regions in which a variation in the number available can make a large contribution to the resistance fluctuations.

The non-reproducible effects at higher currents in the compound type resistor may perhaps be due to very *local* heating effects at internal contacts, though the total power dissipation in the resistors when these effects occur is only of the order of 10^{-5} watt, and any departure from Ohm's law is very slight. The absence of such effects in the film-type resistors may perhaps be ascribed to the much better heat exchange provided by their construction.

4.4. *Temperature Dependence*

From fig. 5 it is clear that Macfarlane's theory is not adequate as it stands to account for the experimental results. The importance of a wide temperature range for such comparisons is also obvious since, for example, over the upper range between room temperature and that of liquid oxygen ($1/T \simeq 3\times 10^{-3}$ to 11×10^{-3}) one might well conclude that there was indeed satisfactory agreement with the predicted linear relationship.

Qualitatively, the shape of the graphs suggests that perhaps an extension of Macfarlane's theory based on *two* activation energies (as in the case of the resistance itself) might account for the observed temperature dependence of the noise. Until, however, a more certain theoretical foundation is established it seems unwise at present to consider the estimation of activation energies from the noise data (particularly in view of the large numerical factor, 8, involved in the exponential index of Macfarlane's formula).

ACKNOWLEDGMENTS

Our thanks are due to Dr. B. V. Rollin, who took considerable interest in the later stages of this work, and also to Dr. J. Wilks and Mr. F. J. Webb for their help with some of the resistance measurements. One of us (I.M.T.) is indebted to the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCES

ABSON, W., HARRIS, E. J., and ROBERTS, W. L., 1946, *Telecommunications Research Establishment Report* (unpublished).

CAMPBELL, R. H., and CHIPMAN, R. A., 1949, *Proc. Inst. Radio Engrs., N.Y.*, **37,** 938.

CHRISTENSEN, C. J., and PEARSON, G. L., 1936, *Bell Syst. Tech. J.*, **15,** 197.

GRISDALE, R. O., PFISTER, A. C., and VAN ROOSBROEK, W., 1951, *Bell Syst. Tech. J.*, **20,** 271.

JOHNSON, J. R., 1925, *Phys. Rev.*, **26,** 71.

MACDONALD, D. K. C., 1949, *Nature, Lond.*, **163,** 637.

MACDONALD, D. K. C., and SARGINSON, K., 1950, *Proc. Roy. Soc.* A, **203,** 223.

MACFARLANE, G. G., 1947, *Proc. Phys. Soc.*, **59,** 366; 1950, *Proc. Phys. Soc.* B, **63,** 807.

MONTGOMERY, H. C., and SHOCKLEY, W., 1950, *Phys. Rev.*, **78,** 646.

MROZOWSKI, S., 1950, *Phys. Rev.*, **77,** 838; 1952, *Ibid.*, **85,** 609.

ROLLIN, B. V., and TEMPLETON, I. M., 1953, *Proc. Phys. Soc.* B, **76,** 259.

SCHOTTKY, W., 1926, *Phys. Rev.*, **28,** 74.

SIMON, F. E., 1936, *Act.* 7 *Int. Congr. Refrig.*, **1,** 367. (See also Daunt, J. G., and Mendelssohn, K., 1938, *Proc. Phys. Soc.*, **50,** 525.)

WALLACE, P. R., 1947, *Phys. Rev.*, **71,** 622, **72,** 258.

The Supercooling of Water

By E. K. BIGG

Department of Meteorology, Imperial College, London

Communicated by B. J. Mason; MS. received 23rd February 1953

Abstract. It is shown that suspension of drops at the interface of two insoluble liquids represents a satisfactory method of studying the supercooling of water. The temperatures of supercooling thus found depend on the volume of the sample and the rate of cooling. The inter-relation between temperature, volume and time is derived theoretically from simple probability considerations and is shown to be consistent with experiment.

---

§1. Introduction

Many researches have been directed to finding how far water can be supercooled since the effect was first reported about two hundred years ago. It is only in recent years, however, that the effect of the volume of the sample on its freezing temperature has been studied systematically, while the effect of the rate of cooling has either been ignored or stated to be unimportant. Heverley (1949) noted a sudden decrease in freezing temperature when the diameters of drops supported on a waxed paper surface were less than 0·4 mm. A thorough statistical study by Dorsch and Hacker (1950) using large numbers of drops on metal surfaces revealed a much broader distribution of freezing temperatures about the mean than Heverley suspected. Levine (1950), in discussing their results, showed that there was a linear relationship between the logarithm of the drop diameter and the mean freezing temperature. He also attempted an explanation on the basis of a distribution of motes of different efficiencies in causing freezing and showed that reasonable agreement could be obtained by a suitable choice of constants. The effect of time was not considered.

Objections to these works are (*a*) the possibility that the solid surfaces used to support the sample of water induced freezing, (*b*) the lack of a satisfactory relation between volume, time and freezing temperature, without appealing to a particular cause of freezing.

It is the aim here to describe how these objections can be overcome.

§2. Experimental Determination of Freezing Temperatures of Drops

(i) *Experimental Techniques*

The various samples of water used in these experiments were stored in plastic bottles from which drops could be squeezed through finely pointed glass nozzles. Drops were suspended either at the interface of two immiscible liquids, one heavier and one lighter than water, or on a hydrophobic film of silicone oil (Drifilm) and covered with liquid paraffin. The drop container was then placed on a mirror near the bottom of a refrigerator having an open top. Light from a concentrated source above the refrigerator illuminated the mirror, and the image of the drops was projected on to a screen by a telescope. Photographically ruled squares on the screen corresponded to 20 micron squares at the drops. The telescope could be rotated to cover a field of view about 20 mm × 5 mm.

The temperature of the drops was measured by a thermocouple placed centrally in the field of view, the hot junction being in melting ice.

The procedure was to map out the location and diameter of each drop and then to mark the temperature at which it froze. Freezing appeared instantaneous, the brilliant image of the lamp in the drop being extinguished as the ice formed.

Slow cooling was accomplished by conduction from the mirror and the cold air surrounding the drop container. It could be hastened by lowering the temperature of the refrigerator, and in this way a standard cooling rate of 0·5° c per minute could be obtained. In the case of very large drops cooling was necessarily slower.

The inaccuracy in the temperature determinations is considered to be of the order of 0·5° c.

It has often been reported that the nature of the surface with which the sample of water is in contact exerts an unpredictable influence on the freezing temperature. Suspension between liquids reduces the difficulties of surface contaminations or the possible effects of force fields. It also eliminates the possibility of communication between neighbouring drops by dendrites thrown from those which have frozen. There are several pairs of suitable common liquids available. Combinations of the following were used: carbon tetrachloride and ethylene dichloride (heavier than water), amyl acetate, toluene and liquid paraffin (lighter than water).

The mean freezing point of each group of 1 mm diameter drops, cooled at about $\frac{1}{2}$°c min^{-1} was found to be within 0·5°c of the overall mean of $-23{\cdot}8$°c. It was found that the mean freezing point of 1 mm diameter drops resting on a hydrophobic film of silicone oil (Drifilm) and cooled at the same rate fell within the same limits.

It was concluded that each of these methods was satisfactory, and in subsequent work the Drifilm under liquid paraffin was used for small drops, while a carbon tetrachloride–liquid paraffin interface was used for large drops. The reason for this choice was that small drops at a liquid interface tend to coalesce while large drops resting on a surface lose their spherical shape.

Experiments were made on water obtained from three sources: (*a*) pure conductivity water (supplied by the National Physical Laboratory) which had been filtered through an ion exchange column, (*b*) once-distilled tap water, from a good laboratory still, (*c*) by direct condensation on to a cold hydrophobic surface from a beaker of hot distilled water.

In each case the mean freezing point of 1 mm diameter drops cooled at $\frac{1}{2}$°c min^{-1} was within 0·5°c of $-23{\cdot}8$°c. The variation of purity within these limits therefore does not greatly affect the freezing temperature.

(ii) *Dependence of Freezing on Volume*

Groups of drops in the diameter range 50 microns to 2·5 cm were studied. Not less than 100 observations were made for each point on fig. 1 with the exception of the point A, which is the average of only 20 observations. For this point the rate of cooling was 2°c h^{-1} and the point A′ represents A corrected for this in the manner to be described in the next section.

The point represented, for example, by diameter 0·65 mm, temperature −25·1°c, was obtained by averaging the freezing temperatures of all drops with diameters in the range 0·60–0·69 mm. If the diameter–temperature relationship is approximately linear over that range and the frequency distribution of freezing symmetrical, the point means that the probability of a drop of 0·65 mm diameter freezing is 0·5. These conditions are very nearly fulfilled.

It will be seen that the logarithm of the diameter is directly proportional to the temperature, a result reported by Levine from the observations of Dorsch and Hacker. The slope of their line is slightly steeper than that of fig. 1, while their freezing points for 1 mm diameter drops are about 5°c higher, with cooling rates 10–30 times as great. Allowance for their higher cooling rates adds another 2–3°c to the discrepancy.

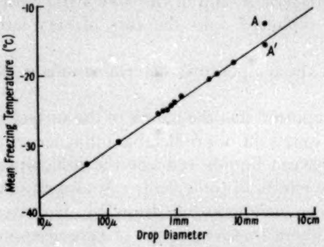


Fig. 1. Mean freezing temperature as a function of drop diameter.

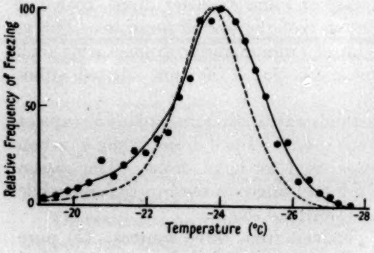


Fig. 2. Relative frequency of freezing as a function of temperature for 1 mm diameter drops.

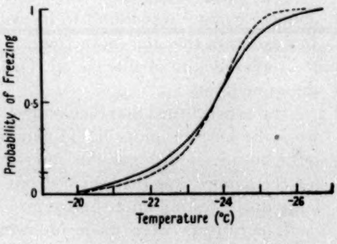


Fig. 3. Probability of freezing as a function of temperature for 1 mm diameter drops.

Once the dependence of freezing point on volume has been established, we may reduce the freezing temperature of any drop to that of a drop of standard size. In this way we may use any of the observations of fig. 1 to obtain a frequency distribution of freezing as a function of temperature. The full lines of fig. 2 embody the results of 1089 observations at temperature intervals of 0·3°c. Only drops with diameters 0·4–1·5 mm were used for this figure as the measurements made on this range were the most accurate.

The standardized integral of this curve gives the probability that a drop will have frozen when the temperature T is reached, and is shown in fig. 3. In figs. 2 and 3 the broken curves are derived theoretically and will be discussed in the next section.

(iii) *Dependence of Freezing on Rate of Cooling*

With the scatter in freezing temperature shown in fig. 2 it is not surprising that investigators using few samples of water have failed to find a dependence of freezing on rate of cooling. The statistical study of Dorsch and Hacker did not reveal an effect, but their cooling rates were varied only from 0·1 to 0·3°C per second. As cooling rates faster than $\frac{1}{2}$°C min^{-1} could lead to difficulty with temperature measurements in the present experiments, a lower rate of cooling was arranged by encasing the drop container in a block of ice and placing it in the refrigerator. The cooling rate could then be made fairly uniform at 3°C per hour. The mean freezing point of 164 drops of 1 mm diameter was then found to be −21·8°C; this was 2·0°C above the mean freezing point with a cooling rate of $\frac{1}{2}$°C min^{-1}, an elevation far too great to be explained by experimental error. The cooling rate therefore has a small but real effect on freezing, deeper supercooling being possible with higher cooling rates.

§ 3. DISCUSSION OF EXPERIMENTAL RESULTS

We have seen that the volume of water, its temperature and the length of time for which it has been cooled all influence the chance that it will freeze. We must now attempt to find a relationship between these which will allow us to check the consistency of the results.

Let us consider an elemental volume of the liquid δV and suppose that the probability of freezing initiating in it in time δt is $W(T_0 - T, \delta V, \delta t)$ where T is the temperature and T_0 is 0°C. For convenience we shall write $T_0 - T$ as T_s, the degree of supercooling.

Now the probability that an event having a chance of occurrence W in a single trial will not have occurred after n trials is $(1-W)^n$. Consequently, in the case in which we are interested, the probability $P(T, V, t)$ that freezing will occur in a volume V and time t is given by

$$(1-P) = (1-W)^{Vt/\delta V \delta t} \quad \text{or} \quad \ln(1-P) = \frac{Vt}{\delta V \delta t} \ln(1-W).$$

If W is very small compared with unity

$$\ln(1-P) = -\frac{VtW}{\delta V \delta t}. \qquad \ldots\ldots(1)$$

In discussing the above experiments where temperature is a function of time it is necessary to proceed somewhat differently because W is a function of temperature. In n successive trials for which the probability of freezing has consecutive values $W_1, W_2, \ldots W_n$, the chance that freezing will not occur is

$$1 - P = (1-W_1)(1-W_2)\ldots(1-W_n).$$

If all the values W_1 to W_n are very much less than unity,

$$\ln(1-P) = -\sum_{i=1}^{n} iW_i. \qquad \ldots\ldots(2)$$

Suppose that cooling takes place at a uniform rate. Then we can write $\alpha t = T_s$.

Now $n = V\,dt/\delta V\,\delta t = V\,dT_s/\alpha\delta V\,\delta t$ and eqn. (2) becomes

$$\ln(1-P) = -\frac{V}{\alpha\delta V\delta t}\int W\,dT_s. \qquad \ldots\ldots(3)$$

Let us write $\int W\,dT_s/\delta V\,\delta t$ as I. Then fig. 1 allows I to be determined as it shows the freezing temperature corresponding to a given drop volume for $P=\frac{1}{2}$. Thus $I = -\alpha \ln \frac{1}{2}/V$.

It can be seen from fig. 1 that $-\ln V = aT_s - b$, where a and b are constants. $a \simeq 6{\cdot}9/6{\cdot}8 \simeq 1{\cdot}0$, $b \simeq 16{\cdot}2$. Then, in the experimental range, $I = 5{\cdot}2\times10^{-10}\exp T_s$. The boundary conditions are that $P=0$ when $T_s=0$ and therefore $I=0$. The correct form is therefore $I = 5{\cdot}2\times10^{-10}(\exp T_s - 1)$. The term $W/\delta V\,\delta t = dI/dT_s$ of eqn. (1) is now $5{\cdot}2\times10^{-10}\exp T_s$. Calculations can therefore be made for steady temperature conditions using this equation. The probability curve of fig. 3 can now be calculated directly from eqn. (3). Also, since $dI/dT_s \simeq I$, $(\partial P/\partial T_s)_V = (1-P)\ln(1-P)$, allowing the broken curve of fig. 2 to be calculated.

The agreement is quite good when one considers that the curve will be broadened by experimental errors. Both experimental and calculated curves show a small degree of asymmetry in the same sense.

The influence of cooling rate can be seen from eqn. (3). The dependence of T on $1/\alpha$ for a given P is the same as the dependence of T on V for a given P. That is, a factor of 1000 in $1/\alpha$ represents 6·78°c, a factor of 10 is equivalent to 2·26°c, and a factor of 2 to 0·55°c. The experimental finding that a factor of 10 decrease in α elevates the mean freezing point by 2·0°c is thus in fair agreement with theory.

Having shown that the results are consistent, and having obtained analytical expressions for the probability of freezing, it is possible to extrapolate to smaller or larger volumes of water.

Extrapolation to cloud drop size (diameter about 10 microns) provides a useful check on this work, for experiments have been made in this range by Cwilong (1947), Fournier d'Albe (1949), Schaefer (1948) and Mason (1952), who report 'critical temperatures' in the range -39 to -42°c.

This is just the range which we should expect from eqn. (1) for a probability of freezing of 1%, if drop diameters varied between 5 and 10μ and times of cooling from 0·1 to 1 second.

It is difficult to compare the results described by fig. 1 with those of other published works on the supercooling of water, either because necessary experimental details are missing, or because of the unknown effects of glass and metal containers. However, if there were valid claims that water had been supercooled to consistently lower temperatures than found here, doubt would be cast on the experimental technique. Consequently, we shall examine freezing temperatures reported by other workers which are frequently referred to in the literature. Beside their experimental values we have placed in brackets the approximate value to be expected from eqn. (3), see table.

It can be seen that of these results the only ones which may be lower than expected are those of Meyer and Pfaff and Smith-Johannsen. The former used small glass phials having a capacity of about 0·6 cm^3. Five of these, containing 'small' quantities of water, were chilled rapidly and two reached -33°c before freezing. Smith-Johannsen also used small quantities of water and cooled them rapidly. These low temperatures could be explained by the small volumes, the rapid cooling, or the finite probability that the freezing temperature was below the mean.

Finally we must refer to the mechanism of freezing, for although the equations give no indication of what this is, they can be compared with one of the current theories.

Author	Volume (cm^3)	Cooling rate	Freezing temp. (°c)	
Meyer and Pfaff (1935)	0·6	rapid	~ -20	(-20)
	smaller quantity	,,	-33	(?)
Tammann and Büchner (1935)	0·4	1°c min^{-1}	-15 to $-18·7$	(-18)
Dorsey (1948)	3	slow	-10 to -20	(-15?)
Smith-Johannsen (1948)	small	rapid	about -20 once at $-38·5$	(?)
Heverley (1949)	5×10^{-4}	1–20°c min^{-1}	-16	(-25)
	5×10^{-7}	,,	-25	(-32)
Dorsch and Hacker (1950)	5×10^{-4}	0·1–0·3° sec^{-1}	-18 av.	(-26)
	5×10^{-7}	,,	-25 av.	(-33)
Brewer and Palmer (1951)	4×10^{-6}	1–20°c min^{-1}	-20	(-31)
	10^{-9}	,,	-30	(-36)

The bulk of experimental evidence favours the view that freezing is initiated by singularities, or motes, in the liquid. If this is so, then the motes must be common constituents of cloud drops and laboratory distilled water. If the singularities are not foreign bodies, but chance aggregates of water molecules into the ice lattice, freezing is said to occur by 'spontaneous nucleation'. Recent work on solidification of liquid metals (see, for example, Turnbull and Fisher 1949) has demonstrated that with suitable experimental conditions freezing occurs as a result of spontaneous nucleation.

Mason (1952) applied the Turnbull and Fisher equations to the problem of nucleation in water, showing that it was not unreasonable to suppose that at -40°c cloud droplets freeze by this process. In Mason's paper the nucleation rate J corresponds to the probability W used here. Now we have found that W is proportional to $\exp T_s$, so that $d(\ln W)/dT_s = \text{const.}$, while Mason shows that, approximately, $d(\ln J)/dT_s = \text{const}/T_s^3$.

The difference can be reconciled if one makes the assumption that the quantity termed the 'interfacial surface energy' is not a constant, but varies as $T_s^{3/2}$. There is no obvious theoretical reason for making such an assumption.

§4. Conclusion

We see then that simple probability considerations are sufficient to resolve the conflicting answers to the question "How far can water be supercooled?". We have seen also that it does not require elaborate apparatus or difficult techniques to obtain consistent data on supercooling.

The more fundamental problem of why supercooling terminates remains to be solved.

Acknowledgments

I wish to thank Messrs. B. J. Mason and D. J. Moore of the Imperial College for valuable suggestions, some of which were incorporated in this paper.

REFERENCES

Brewer, A. W., and **Palmer**, H. P., 1951, *Proc. Phys. Soc.* B, **64,** 772.
Cwilong, B. M., 1947, *Proc. Roy. Soc.* A, **190,** 137.
Dorsch, R. G., and **Hacker**, P., 1950, *N.A.C.A. Tech. Notes* 2142.
Dorsey, N. E., 1948, *Trans. Amer. Phil. Soc.*, **38,** 248.
Fournier d'Albe, E. M., 1949, *Quart. J. Roy. Met. Soc.*, **75,** 1.
Heverley, J. R., 1949, *Trans. Amer. Geophys. Un.*, **30,** 205.
Levine, J., 1950, *N.A.C.A. Tech. Notes* 2234.
Mason, B. J., 1952, *Quart. J. Roy. Met. Soc.*, **78,** 22 .
Meyer, J., and **Pfaff**, W., 1935, *Z. Anorg. Chem.*, **224,** 305.
Schaefer, V. J., 1948, *Bull. Amer. Met. Soc.*, **29,** 175.
Smith-Johannsen, R., 1948, *Science*, **108,** 652.
Tammann, G., and **Büchner**, A., 1935, *Z. Anorg. Chem.*, **222,** 371.
Turnbull, D., and **Fisher**, J. C., 1949, *J. Chem. Phys.*, **17,** 71.

The Theory of the Propagation of Plane Sound Waves in Tubes

By D. E. WESTON*
Physics Department, Imperial College, London

Communicated by R. W. B. Stephens ; *MS. received* 26*th March* 1953

Abstract. The propagation of plane sound waves in gases in tubes can be divided into three main types, depending on the radius and frequency involved. These types are described as 'narrow' tube, 'wide' tube and 'very wide' tube propagation. The phase velocity, attenuation and cross section profile of particle velocity etc. are investigated theoretically, and their inter-relation pointed out. The factors affecting the validity of Kirchhoff's formulae are considered, and the theory is applied to some recent work.

---

§ 1. Introduction

This paper is concerned with the theory for the propagation of sound waves of the principal mode in gases contained in cylindrical tubes of a large range of diameters. In particular the manner in which viscosity and thermal conductivity affect the motion is discussed for continuous waves.

Helmholtz (1863) was the first to investigate the problem, and quoted the formula for the velocity decrease due to viscosity. In his dust-tube experiment Kundt (1868) obtained fair agreement with the Helmholtz formula, but found the effect larger than predicted. He suggested heat conduction as a further cause, stimulating Kirchhoff (1868) to present the complete theory for the 'wide' tube, which is the most important case.

Kirchhoff's formulae are:

Phase velocity of sound:

$$c' = c[1 - \gamma'/r_{\mathrm{w}}(2\omega)^{1/2}] = 3{\cdot}434 \times 10^4\{1 - 0{\cdot}162/(r_{\mathrm{w}}\sqrt{f})\}\,\mathrm{cm\ sec^{-1}}. \quad ..(1)$$

Amplitude attenuation constant:

$$m' = (\gamma'/cr_{\mathrm{w}})(\omega/2)^{1/2} = 2{\cdot}964 \times 10^{-5}\sqrt{f}/r_{\mathrm{w}}\ \mathrm{cm^{-1}}. \quad ......(2)$$

The numerical values here and elsewhere are for air at 20°c and 76 cm Hg; ω is the pulsatance, f the frequency, r_{w} the tube radius and r the general radius vector, and the following refer to the gaseous medium:

$$\left.\begin{array}{l} c = \text{Laplacian adiabatic velocity of sound,}\\ \gamma' = \sqrt{\nu} + (\gamma - 1)(\nu'/\gamma)^{1/2} = 0{\cdot}574 \text{ c.g.s. units} = \text{Kirchhoff's}\\ \qquad\text{constant,}\\ \gamma = \text{ratio of the principal specific heats,}\\ \nu = \mu/\rho = \text{kinematic viscosity,}\\ \nu' = k/\rho c_{\mathrm{v}} = \text{thermal diffusivity or thermometric}\\ \qquad\text{conductivity,}\\ \mu = \text{shear viscosity coefficient,}\\ k = \text{thermal conductivity,}\\ \rho = \text{density,}\\ c_{\mathrm{v}} = \text{specific heat at constant volume.} \end{array}\right\} \quad ......(3)$$

* Now at the Admiralty Research Laboratory, Teddington.

The author has made a critical examination of Kirchhoff's formulae in an effort to settle the controversy arising from the disagreement of many experimental results with theory. Using some results of Kirchhoff it may be shown that there are three main types of motion, the conditions for which are defined later (see table 1 and fig. 1):

(*a*) 'Narrow' tube. At very small radii and low frequencies the motion is isothermal and governed largely by viscous forces.

(*b*) 'Wide' tube. The sound energy is diffused evenly over the tube cross section, and the adiabatic motion approximates to that in an unbounded gas. There is, however, a narrow boundary layer at the wall, analogous to the Prandtl layer in continuous flow, where viscous and heat conduction processes lead to an energy loss.

(*c*) 'Very wide' tube. This differs from the 'wide' tube mainly in that the sound energy is concentrated near the walls.

Before giving the more formal analysis it is desirable to consider two alternative approaches to the problem—the resistance concept for 'narrow' tubes and the boundary layer idea for 'wide' tubes. Besides affording a clearer physical picture, these cover aspects additional to the main theory.

§ 2. THE RESISTANCE CONCEPT FOR NARROW TUBES

When the tube radius is much less than the boundary layer thickness a simple approach due to Lamb (1898) is valid. The motion is isothermal and viscous forces predominate over inertial forces. If u is the mean velocity over the cross section

$$Ru = -\partial p/\partial x \qquad \ldots\ldots(4)$$

where R is a coefficient of resistance and $\partial p/\partial x$ the pressure gradient along the tube axis. This equation leads to the result

$$m^2 = i\omega R/P \qquad \ldots\ldots(5)$$

where m is the complex propagation constant and P the ambient pressure.

The value of R is usually calculated assuming that Poiseuille's law is applicable, viz.

$$R = 8\mu/r_w{}^2. \qquad \ldots\ldots(6)$$

Thus for air, velocity $c' = \frac{1}{2}cr_w(\omega/\gamma\nu)^{1/2} = 9{\cdot}364 \times 10^4 r_w\sqrt{f}\ \mathrm{cm\,sec^{-1}}$ ......(7)

attenuation $m' = 2(\gamma\nu\omega)^{1/2}/cr_w = 6{\cdot}701 \times 10^{-5}\sqrt{f}/r_w\ \mathrm{cm^{-1}}$. ......(8)

For the narrower tubes R may be modified for the effect of slip, and for very narrow tubes it may be found from Knudsen's formula for molecular streaming (see eqn. (72)).

§ 3. THE BOUNDARY LAYER CONCEPT FOR WIDE TUBES

(i) *Viscosity and Temperature Waves*

When a fluid is vibrating parallel to a plane solid surface the particle velocity u is given by (Stokes 1851)

$$u = A\,e^{i\omega t}[1 - e^{-\alpha b(1+i)}] \qquad \ldots\ldots(9)$$

where $\alpha = (\omega/2\nu)^{1/2}$ is the viscous 'wave constant', and b is a coordinate measured normal to the solid surface. The first term represents the acoustic wave and the second term a 'viscous' wave.

It may easily be shown that (see e.g. Sexl 1930) the amplitude u_0 of the particle velocity has a principal maximum equal to $1{\cdot}067A$ at a distance from

the surface
$$\Delta_\mu = 2{\cdot}2838(2\nu/\omega)^{1/2} = 0{\cdot}500/\sqrt{f}\ \text{cm for air,} \qquad \ldots\ldots(10)$$
which is here defined as the boundary layer thickness.

Similarly it may be shown that the excess temperature (cf. Kirchhoff 1868, Rayleigh 1896, p. 322, Ballantine 1932)
$$\theta = B\,\mathrm{e}^{i\omega t}[1 - \mathrm{e}^{-\beta b(1+i)}] \qquad \ldots\ldots(11)$$
where $\beta = (\omega\gamma/2\nu')^{1/2}$ is the thermal 'wave constant'. The principal maximum of θ_0 occurs at
$$\Delta_k = 2{\cdot}2838(2\nu'/\omega\gamma)^{1/2} = 0{\cdot}596/\sqrt{f}\ \text{cm for air.} \qquad \ldots\ldots(12)$$

It may be noted that $\Delta_\mu \sim \Delta_k \sim (\lambda\lambda')^{1/2}$, the geometric mean of the wavelength and mean free path.

The shape of the wave front may also be calculated from eqn. (9). Thus
$$u = u_0 \cos(\omega t - m''x + \phi) \qquad \ldots\ldots(13)$$
where
$$\tan\phi = \mathrm{e}^{-\alpha b}\sin\alpha b/(1 - \mathrm{e}^{-\alpha b}\cos\alpha b) \qquad \ldots\ldots(14)$$
ϕ has a maximum of $\pi/2$ at the wall and a principal minimum of $\phi = -0{\cdot}00438\pi$, corresponding to an appreciable forward displacement (about 0·2% wavelength) at
$$b = 3{\cdot}9408(2\nu/\omega)^{1/2}. \qquad \ldots\ldots(15)$$

Figure 3 shows the velocity amplitude and wave front when the above theory is applied to a tube. The maxima due to viscous waves have been found experimentally by Richardson (1928), Carriere (1929), Richardson and Tyler (1929), and others.

(ii) *Velocity and Attenuation*

Let the perimeter and cross-sectional area of the tube be E and S, and the excess pressure outside the layer
$$p = p_0 \exp(i\omega t - mx). \qquad \ldots\ldots(16)$$
The variable component of the particle velocity near the wall is
$$u = -(cp_0/\gamma P)\exp\{-\alpha b(1+i) + i\omega t - mx\}. \qquad \ldots\ldots(17)$$
From Newton's law the force on a cylindrical gas column of length δx is
$$E\delta x\mu\left(\frac{\partial u}{\partial b}\right)_{b=0} = \frac{Ep_0\delta x}{c}\left(\frac{\nu\omega}{2}\right)^{1/2}(1+i)\exp(i\omega t - mx). \qquad \ldots\ldots(18)$$
The associated pressure is found to be
$$(E/S)(\nu/2\omega)^{1/2}(i-1)p. \qquad \ldots\ldots(19)$$
The variable component of the temperature excess near the wall is
$$\theta = -\{p_0(\gamma-1)\Theta/\gamma P\}\exp\{-\beta b(1+i) + i\omega t - mx\} \qquad \ldots\ldots(20)$$
where Θ is the ambient temperature. The pressure arising is
$$-\{p_0(\gamma-1)/\gamma\}\exp\{-\beta b(1+i) + i\omega t - mx\}. \qquad \ldots\ldots(21)$$
This corresponds to a pressure, averaged over the cross section,
$$\frac{1}{S}\int_0^\infty -\frac{Ep_0(\gamma-1)}{\gamma}\exp\{-\beta b(1+i) + i\omega t - mx\}\,db = \frac{E}{S}\,\frac{\gamma-1}{\gamma^{3/2}}\left(\frac{\nu'}{2\omega}\right)^{1/2}(i-1)p. \qquad \ldots\ldots(22)$$

If the total *fractional* change in p due to heat conduction is ϵ there will be a corresponding fractional change ϵ in the temperature excess in the bulk of the tube. This produces part of the change $\epsilon(\gamma-1)/\gamma$ in p, from the constant volume relation. Thus ϵ has two components:

$$\epsilon=\frac{E}{S}\frac{\gamma-1}{\gamma^{3/2}}\left(\frac{\nu'}{2\omega}\right)^{1/2}(i-1)+\frac{\gamma-1}{\gamma}\epsilon. \qquad \ldots\ldots(23)$$

Solving eqn. (23)

$$\epsilon=\frac{E}{S}\frac{\gamma-1}{\sqrt{\gamma}}\left(\frac{\nu'}{2\omega}\right)^{1/2}(i-1). \qquad \ldots\ldots(24)$$

The extra pressure due to both viscosity and heat conduction is

$$\{E\gamma'/S(2\omega)^{1/2}\}(i-1)p \qquad \ldots\ldots(25)$$

and the wave equation is modified to

$$\frac{\partial^2\xi}{\partial t^2}=c^2\frac{\partial^2\xi}{\partial x^2}\left[1+\frac{E}{S}\frac{\gamma'}{(2\omega)^{1/2}}(i-1)\right]. \qquad \ldots\ldots(26)$$

The phase velocity and attenuation become

$$c'=c\{1-E\gamma'/2S(2\omega)^{1/2}\} \qquad \ldots\ldots(27)$$

$$m'=(E\gamma'/2Sc)(\omega/2)^{1/2}. \qquad \ldots\ldots(28)$$

These reduce to (1) and (2) for the circular tube.

A version of the above theory for the effect of viscosity has been given by Rayleigh (1896) and others, but the heat conduction layer has not received such full attention (see however Ballantine 1932, Nielsen 1949 a, b, Konstantinov 1939, Cremer 1948).

§ 4. THE MORE COMPLETE THEORY

(i) *General Equations*

Kirchhoff (1868, see alternatively Rayleigh 1896), when analysing propagation in a tube, started from differential equations modified for both viscosity and thermal conduction. He derived expressions showing the variations across the tube of the particle velocity u parallel to the axis, the radial particle velocity q, and temperature excess θ. These, however, involve constants A, A_1, A_2 of values to be determined by the boundary conditions,

$$\left.\begin{aligned}u&=AQ-A_1m(h/\lambda_1-\nu')Q_1-A_2m(h/\lambda_2-\nu')Q_2\\ q&=\frac{-Am}{h/\nu-m^2}\frac{dQ}{dr}-A_1\left(\frac{h}{\lambda_1}-\nu'\right)\frac{dQ_1}{dr}-A_2\left(\frac{h}{\lambda_2}-\nu'\right)\frac{dQ_2}{dr}\\ \theta'&=A_1Q_1+A_2Q_2\end{aligned}\right\} \qquad \ldots\ldots(29)$$

where $\theta'=\theta/(\gamma-1)\Theta$, $h=i\omega$ and

$$Q=J_0\{r(m^2-h/\nu)^{1/2}\} \quad Q_1=J_0\{r(m^2-\lambda_1)^{1/2}\} \quad Q_2=J_0\{r(m^2-\lambda_2)^{1/2}\} \qquad \ldots\ldots(30)$$

λ_1 and λ_2 are the small and large roots of

$$h^2-\{c^2+h(\nu+\nu''+\nu')\}\lambda+(\nu'/h)\{c^2/\gamma+h(\nu+\nu'')\}\lambda^2=0 \qquad \ldots\ldots(31)$$

ν'' is a constant which according to Stokes equals $\nu/3$.

At the walls of a rigid conducting tube, where $r=r_w$,

$$u=q=\theta'=0. \qquad \ldots\ldots(32)$$

The vanishing of the determinant (29) gives an equation for the propagation constant m used by Kirchhoff,

$$\frac{m^2h}{h/\nu-m^2}\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right)\frac{d\ln Q}{dr_w}+\left(\frac{h}{\lambda_1}-\nu'\right)\frac{d\ln Q_1}{dr_w}-\left(\frac{h}{\lambda_2}-\nu'\right)\frac{d\ln Q_2}{dr_w}=0. \qquad \ldots(33)$$

The profiles, i.e. distributions of particle velocity etc. across the tube, may be expressed conveniently in terms of another constant B. The suffix w to a quantity denotes its value at the wall.

$$\left.\begin{aligned} u &= \left[-h\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right)Q_{1\mathrm{w}}Q_{2\mathrm{w}}Q+\left(\frac{h}{\lambda_1}-\nu'\right)Q_{\mathrm{w}}Q_{2\mathrm{w}}Q_1\right.\\ &\qquad\left.+\left(\frac{h}{\lambda_2}-\nu'\right)Q_{\mathrm{w}}Q_{1\mathrm{w}}Q_2\right]mB\\ q &= \left[\frac{m^2h}{h/\nu-m^2}\left(\frac{1}{\lambda_1}-\frac{1}{\lambda_2}\right)Q_{1\mathrm{w}}Q_{2\mathrm{w}}\frac{dQ}{dr}+\left(\frac{h}{\lambda_1}-\nu'\right)Q_{\mathrm{w}}Q_{2\mathrm{w}}\frac{dQ_1}{dr}\right.\\ &\qquad\left.-\left(\frac{h}{\lambda_2}-\nu'\right)Q_{\mathrm{w}}Q_{1\mathrm{w}}\frac{dQ_2}{dr}\right]B\\ \theta' &= [-Q_{2\mathrm{w}}Q_1+Q_{1\mathrm{w}}Q_2]Q_{\mathrm{w}}B. \end{aligned}\right\}\quad\ldots\ldots(34)$$

The form of solution depends on whether the moduli of Q, Q_1, Q_2 are small or large (see table 1 and fig. 1), and the various regions are now considered.

Table 1. The Various Propagation Regions (at 20°c and 76 cm Hg)

'Very small' implies one term, and 'small' several terms in eqn. (38); 'large' implies several terms, and 'very large' one term in eqn. (43).

Type of propagation	Magnitude of modulus			Discriminant
	Q	Q_1	Q_2	
Narrow tube	very small	very small	very small	
Transition, narrow–wide	small	very small	small	$r_{\mathrm{w}}(\omega/\nu)^{1/2}\sim 2r_{\mathrm{w}}\sqrt{f}$
Transition, wide–narrow	large	very small	large	
Wide tube	very large	very small	very large	
Transition, wide–very wide	very large	small	very large	$r_{\mathrm{w}}\omega^{3/2}\gamma'/c^2\sim 10^{-8}r_{\mathrm{w}}f^{3/2}$
Transition, very wide–wide	very large	large	very large	
Very wide tube	very large	very large	very large	

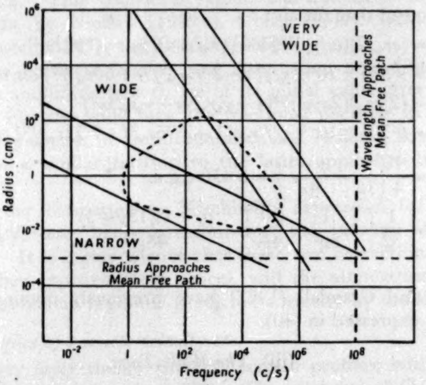


Fig. 1. The various propagation regions in air at 20°c and 76 cm Hg showing where the correction terms in the formulae for the attenuation are less than 1%. Reported experimental work lies within the dotted line.

(ii) *The Narrow Tube*

Cross section profiles of particle velocity etc.

The original treatment for waves in a narrow tube was given by Rayleigh (1896). First approximations from (31) are

$$\lambda_1 = h^2/c^2, \qquad \lambda_2 = h\gamma/\nu'. \qquad \ldots\ldots(35)$$

An approximation valid at all but the highest frequencies is $h\nu/c^2 \ll 1$, and for the narrow tube $(\gamma'/r_w\sqrt{h} \gg 1)$

$$m^2 = 8\nu\gamma h/c^2 r_w^{\,2}. \qquad \ldots\ldots(36)$$

Retaining only the leading terms the profiles may be calculated from (34)

$$\left.\begin{aligned} u &= (Bc/r_w)(\gamma h/2\nu)^{1/2}(r_w^{\,2}-r^2) \\ q &= -(Bhr/r_w^{\,2})(2\gamma-1)(r_w^{\,2}-r^2) \\ \theta' &= -(Bh\gamma/4\nu')(r_w^{\,2}-r^2). \end{aligned}\right\} \qquad \ldots\ldots(37)$$

The axial distribution is parabolic, as for continuous flow in tubes. Sexl (1928, 1930) also finds that his general expression, allowing for viscosity only, reduces to a Poiseuille type flow for narrow tubes. Note that $q/u \sim (\omega\nu)^{1/2}/c$, i.e. the radial is much less than the axial particle velocity. Compared with propagation in an unbounded gas θ'/u is multiplied by the small factor $r_w(\omega/\nu)^{1/2}$, so the flow may be called isothermal.

Transition to the wide tube.

Starting from the two extremes it is possible to derive expressions applicable to the transition region between the narrow and wide tube. For the narrow tube approach, it may be shown, on developing the power expansion of a zero order Bessel function for small x, that

$$\frac{d \ln J_0(x)}{dx} = -\frac{x}{2}\left[1 + \frac{x^2}{8} + \frac{x^4}{48} + \frac{11x^6}{3072}\cdots\right]. \qquad \ldots\ldots(38)$$

By making approximations in eqn. (33) as in table 1, but otherwise following Rayleigh, it is found that for air:

$$\begin{aligned} &\text{velocity} & c' &= (cr_w/2)(\omega/\nu\gamma)^{1/2}[1-\gamma_1 r_w^{\,2}\omega + (\gamma_2+\gamma_1^{\,2})r_w^{\,4}\omega^2] \\ && &= 9{\cdot}364\times 10^4 r_w\sqrt{f}[1-2{\cdot}95 r_w^{\,2}f + 1{\cdot}60 r_w^{\,4}f^2]\ \text{cm sec}^{-1} \qquad \ldots(39) \\ &\text{attenuation} & m' &= (2/cr_w)(2\nu\gamma\omega)^{1/2}[1-\gamma_1 r_w^{\,2}\omega - \gamma_2 r_w^{\,4}\omega^2] \\ && &= 6{\cdot}701\times 10^{-5}(\sqrt{f}/r_w)[1-2{\cdot}95 r_w^{\,2}f + 7{\cdot}11 r_w^{\,4}f^2]\ \text{cm}^{-1} \qquad \ldots(40) \end{aligned}$$

$$\left.\begin{aligned} \text{where} \quad \gamma_1 &= \frac{1}{4}\left\{\frac{1}{3\nu} - \frac{\gamma-1}{4\nu'}\right\} = 0{\cdot}470 \text{ c.g.s.} \\ \gamma_2 &= \frac{1}{32}\left\{-\frac{1}{8\nu^2} - \frac{\gamma-1}{4\nu\nu'} + \frac{(\gamma-1)(13\gamma+3)}{48\nu'^2}\right\} = -0{\cdot}180 \text{ c.g.s.} \end{aligned}\right\} \qquad \ldots\ldots(41)$$

Kosten (1949) and Crandall (1927) have previously obtained a part of the correction term expressed in (40).

(iii) *The Wide Tube*

Transition to the narrow tube.

When the boundary layer thickness Δ becomes comparable with the tube radius Kirchhoff's assumptions begin to break down: in particular the moduli of Q and Q_2 are no longer large. The discriminant between the two cases of adiabatic and isothermal flow is thus

$$r_w(\omega/\nu)^{1/2} \propto r_w/\Delta \propto \text{moduli of } Q,\ Q_2 \sim r_w\sqrt{f}. \qquad \ldots\ldots(42)$$

For x large

$$\frac{d \ln J_0(x)}{dx} = -i\left(1+\frac{1}{8x^2}-\frac{25}{128x^4}\cdots\right)-\left(\frac{1}{2x}-\frac{1}{8x^3}\cdots\right) \qquad \ldots\ldots(43)$$

where for the conditions considered it has been possible to approximate $\tan(x-\frac{1}{4}\pi)=i$. By solving eqn. (33) with approximations as in table 1, but otherwise following Kirchhoff, it is found that for air:

$$\text{velocity} \qquad c' = c\left[1-\frac{\gamma'}{r_w(2\omega)^{1/2}}+\frac{\gamma'^2}{2r_w^2\omega}-\frac{\gamma'''+\gamma'^3/2}{r_w^3(2\omega^3)^{1/2}}\right]$$

$$= 3{\cdot}434\times 10^4\left[1-\frac{0{\cdot}162}{r_w\sqrt{f}}+\frac{0{\cdot}0262}{r_w^2 f}-\frac{0{\cdot}0124}{r_w^3 f^{3/2}}\right] \text{cm sec}^{-1} \quad \ldots(44)$$

$$\text{attenuation } m' = \frac{\gamma'}{r_w c}\left(\frac{\omega}{2}\right)^{1/2}+\frac{\gamma''}{r_w^2 c}+\frac{\gamma'''}{r_w^3 c(2\omega)^{1/2}}$$

$$= 10^{-5}\left[2{\cdot}964\,\frac{\sqrt{f}}{r_w}+\frac{0{\cdot}473}{r_w^2}+\frac{0{\cdot}146}{r_w^3\sqrt{f}}\right]\text{cm}^{-1} \qquad \ldots\ldots(45)$$

$$\text{where} \qquad \left.\begin{aligned} \gamma'' &= \nu+\frac{(\gamma-1)}{\sqrt{\gamma}}(\nu\nu')^{1/2}-\frac{(\gamma-1)}{2}\nu' = 0{\cdot}162 \text{ c.g.s.} \\ \gamma''' &= \frac{15}{8}\nu^{3/2}+\frac{4(\gamma-1)}{\sqrt{\gamma}}\nu\nu'^{1/2}+\frac{3(\gamma-1)(\gamma-2)}{2\gamma}\nu^{1/2}\nu' \\ &\quad +\frac{(\gamma-1)(4\gamma^2-12\gamma+7)}{8\gamma^{3/2}}\nu'^{3/2} = 0{\cdot}182 \text{ c.g.s.} \end{aligned}\right\} \qquad \ldots\ldots(46)$$

The two sets of approximate expressions for velocity and attenuation, approaching from the wide and narrow tube limiting cases, have been plotted together in fig. **2.** This transition region has also been investigated, using different methods, by Nielsen (1949a), Golay (1947), Daniels (1947, 1950), Zwikker and Kosten (1949), and Iberall (1950).

The relations between velocity and attenuation are of interest. For the narrow tube the amplitude falls to 1/e of its initial value after a phase advance of one radian. For the wide tube the fractional decrease in velocity is equal to the fractional decrease in amplitude after one radian. Bradfield (1951) has pointed out that a similar relation to the latter applies to other propagation processes.

The first order correction to Kirchhoff's expression for the attenuation is $1+0{\cdot}160/r_w\sqrt{f}$, and his expression for the velocity decrease involves $1-0{\cdot}162/r_w\sqrt{f}$. It is a coincidence that these two corrections are numerically close, so that they approximately cancel, and the attenuation per wavelength just inside the transition region is predicted correctly by the simple formula.

Cross section profiles of particle velocity etc.

The boundary layer theory predicts certain profiles, which are confirmed on developing the exact equations (34). The modulus of Q_1 is assumed small and those of Q and Q_2 large, and the relation

$$m^2 = (h^2/c^2)(1+2\gamma'/r_w\sqrt{h}) \qquad \ldots\ldots(47)$$

is used. Introducing a new constant B' the following expressions are obtained when near the wall:

$$\left.\begin{aligned} u &= B'[1-(r/r_{\mathrm{w}})^{1/2}\exp\{-\alpha b(1+i)\}] \\ q &= \frac{B'\sqrt{\omega}\,\mathrm{e}^{-i3\pi/4}}{c}\left[\frac{r}{r_{\mathrm{w}}}\gamma' - \left(\frac{r}{r_{\mathrm{w}}}\nu\right)^{1/2}\exp\{-\alpha b(1+i)\}\right. \\ &\qquad\qquad \left. -(\gamma-1)\left(\frac{r\nu'}{r_{\mathrm{w}}\gamma}\right)^{1/2}\exp\{-\beta b(1+i)\}\right] \\ \theta' &= (B'/c)[1-(r/r_{\mathrm{w}})^{1/2}\exp\{-\beta b(1+i)\}] \end{aligned}\right\} \quad \ldots\ldots(48)$$

The amplitudes of these quantities are plotted in fig. 3. It may be emphasized here that the enhanced particle velocity in the annulus shown in fig. 3 is a consequence of Kirchhoff's theory, and cannot be used to explain an experimental attenuation greater than the theoretical. The radial velocity has two components of the same order of magnitude travelling inwards from the wall, one with the velocity of viscosity waves and one with that of temperature waves. For continuous laminar flow it may be shown (Goldstein 1938, p. 51) that there is a velocity q in the layer normal to the surface such that $q/u \sim \Delta/l$ where l is the distance traversed from the leading edge of a plate. It may be seen from (48) that this is also true for the acoustic layer, when l is the wavelength in the gas. There is not always a motion normal to the surface: for example q is zero in continuous or oscillatory rotational motion of a cylinder about its axis, since the phase does not vary over the surface. It may be inferred from an expression

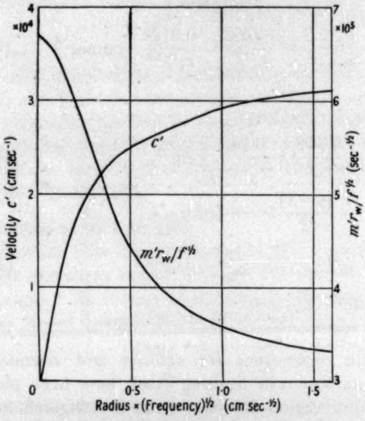


Fig. 2. Theoretical velocity c' and attenuation m' in a tube as a function of $r_{\mathrm{w}}\sqrt{f}$ for air at 20°c, 76 cm Hg.

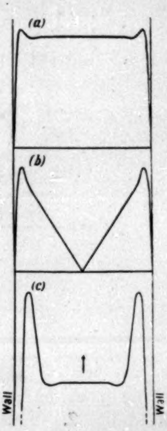


Fig. 3. Cross section profiles in a wide tube. (*a*) amplitude of axial particle velocity or temperature fluctuation, (*b*) amplitude of radial particle velocity, (*c*) wave front or equiphase surface.

due to Kirchhoff (see Rayleigh 1896, p. 322) that there is a particle velocity accompanying a temperature wave such that

$$q/\theta' = (\gamma - 1)(\omega \nu'/\gamma)^{1/2} \exp(-i3\pi/4) \qquad \ldots\ldots(49)$$

agreeing with (48). Also $u/\theta' = c$ outside the layer, which is the normal adiabatic relation.

Near the axis where the moduli of Q and Q_2 are small the above analysis is not applicable. Modifying the analysis, only the first terms in (48) remain; this leads for example to the expression which satisfies the condition that the radial velocity falls to zero at the axis:

$$q = B' \sqrt{\omega}\, e^{-i3\pi/4} \gamma' r / c r_w. \qquad \ldots\ldots(50)$$

The significance of the radial particle velocity.

In the boundary layer theory for the velocity and attenuation the radial velocity was not considered, and this was permissible because the integral over the cross section of the pressure due to the velocity q must vanish.

Rewriting the temperature wave term in q from (48)

$$\begin{aligned} q &= (\gamma - 1)(\omega \nu'/\gamma)^{1/2}\, e^{-i3\pi/4} \theta' \\ &= -\{p_0(\gamma - 1)(\omega \nu')^{1/2}/P\gamma^{3/2}\} \exp\{-i3\pi/4 - \beta b(1+i) + i\omega t - mx\}. \qquad \ldots\ldots(51) \end{aligned}$$

The corresponding pressure excess

$$P \frac{d}{db} \int q\, dt = p_0\{(\gamma - 1)/\gamma\} \exp\{-\beta b(1+i) + i\omega t - mx\}. \qquad \ldots\ldots(52)$$

This is equal in magnitude and opposite in sign to (21) so that altogether the temperature wave produces no pressure in the boundary layer. Similarly no pressure is developed in the layer by the viscosity wave. It should therefore be possible to calculate the extra pressure due to viscosity and heat conduction from the behaviour of q outside the layer, where from (50)

$$q = (p\gamma' r / P\gamma r_w) \sqrt{\omega}\, e^{-i3\pi/4}, \qquad \ldots\ldots(53)$$

and the corresponding displacement

$$\xi = (p\gamma' r / i\omega P\gamma r_w) \sqrt{\omega}\, e^{i\pi/4}. \qquad \ldots\ldots(54)$$

The corresponding pressure may be found using the adiabatic relation, since outside the layer q is associated with what is, essentially, an ordinary sound wave. The extra pressure is

$$-\gamma P[\partial\xi/\partial r + \xi/r] = \{2\gamma'/r_w(2\omega)^{1/2}\}(i-1)p \qquad \ldots\ldots(55)$$

which agrees with (25). Thus the radial particle velocity has the effect of cancelling the extra forces at the boundary and distributing them evenly over the cross section.

Transition to the very wide tube.

When the tube is so wide that the free gas attenuation begins to become comparable with the tube attenuation, Kirchhoff's assumption that the modulus of Q_1 is very small breaks down. Substituting in (33) as did Kirchhoff, but without developing Q_1, we find

$$d \ln Q_1 / dr_w = -h^{3/2}\gamma'/c^2. \qquad \ldots\ldots(56)$$

Using the first two terms of (38), (see table 1)

$$-(r_w/2)(m^2 - \lambda_1)[1 + r_w^2(m^2 - \lambda_1)/8] = -h^{3/2}\gamma'/c^2. \qquad \ldots\ldots(57)$$

Hence $$m^2-\lambda_1=(2h^{3/2}\gamma'/r_{\mathrm{w}}c^2)-(h^3\gamma'^2/2c^4). \qquad \ldots\ldots(58)$$

A second approximation to λ_1 is, from (31),

$$\lambda_1=(h^2/c^2)[1-(h/c^2)\{\nu+\nu''+(\gamma-1)\nu'/\gamma\}]. \qquad \ldots\ldots(59)$$

The velocity and attenuation may be found from the complex propagation constant as before. The velocity is still predicted correctly by Kirchhoff's formula (1), but the attenuation is

$$\begin{aligned} m' &= \frac{\gamma'}{r_{\mathrm{w}}c}\left(\frac{\omega}{2}\right)^{1/2}+\frac{\omega^2}{2c^3}\left\{\nu+\nu''+\frac{\gamma-1}{\gamma}\nu'\right\}+\frac{\omega^2\gamma'^2}{4c^3} \\ &= 2{\cdot}964\times10^{-5}\sqrt{f}/r_{\mathrm{w}}+1{\cdot}403\times10^{-13}f^2+0{\cdot}806\times10^{-13}f^2\ \mathrm{cm}^{-1}\text{ for air}. \qquad \ldots(60)\end{aligned}$$

The first term is the ordinary tube absorption, the second the ordinary free gas attenuation, and the third arises because the sound energy is not uniformly distributed over the tube but tends to concentrate near the walls. Bogert (1950) considering viscosity only and using an approximate method, obtained a similar result as far as the second term.

The profiles outside the boundary layer are

$$\left.\begin{aligned} u&=B'[1+\omega^{3/2}\,\mathrm{e}^{-i\pi/4}\,r^2/2c^2r_{\mathrm{w}}] \\ q&=(B'\gamma'\sqrt{\omega}\,\mathrm{e}^{-i3\pi/4}\,r/cr_{\mathrm{w}})[1+\omega^{3/2}\,\mathrm{e}^{-i\pi/4}\,r_{\mathrm{w}}/4c^2] \\ \theta'&=(B'/c)[1+\omega^{3/2}\,\mathrm{e}^{-i\pi/4}\,r^2/2c^2r_{\mathrm{w}}]\end{aligned}\right\}. \qquad \ldots\ldots(61)$$

The axial velocity and temperature excess profiles have a parabolic curvature, being a minimum at the axis. The wave front is also parabolic, and convex towards $+x$.

(iv) *The Very Wide Tube*

Velocity and attenuation.

In the limit as the radius increases the modulus of Q_1 becomes large so that formula (43) must be used, and (56) becomes

$$(\lambda_1-m^2)^{1/2}-1/2r_{\mathrm{w}}=-h^{3/2}\gamma'/c^2. \qquad \ldots\ldots(62)$$

This leads to

$$c'=c\left\{1-\frac{\gamma'}{2r_{\mathrm{w}}(2\omega)^{1/2}}\right\}=3{\cdot}434\times10^4\left\{1-\frac{0{\cdot}081}{r_{\mathrm{w}}\sqrt{f}}\right\}\mathrm{cm\,sec}^{-1}\text{ for air} \qquad \ldots\ldots(63)$$

$$\begin{aligned} m' &= \frac{\gamma'}{2cr_{\mathrm{w}}}\left(\frac{\omega}{2}\right)^{1/2}+\frac{\omega^2}{2c^3}\left\{\nu+\nu''+\frac{\gamma-1}{\gamma}\nu'\right\}+\frac{\omega^2\gamma'^2}{2c^3} \\ &= 1{\cdot}482\times10^{-5}\sqrt{f}/r_{\mathrm{w}}+1{\cdot}403\times10^{-13}f^2+1{\cdot}612\times10^{-13}f^2\ \mathrm{cm}^{-1}\text{ for air}. \qquad \ldots(64)\end{aligned}$$

The velocity change and first absorption term have half their wide tube values. The boundary layer derivation given above is not valid because phase and intensity vary considerably over the cross section. The first attenuation term is now only a minor correction, the second is unchanged, and the last is doubled in comparison with (60).

Cross section profiles of particle velocity etc.

The principal terms in the profiles (34) outside the boundary layer and away from the axis are (written in terms of a new constant B'')

$$\left.\begin{aligned} u&=(B''/\sqrt{r})\exp\{r(1-i)\gamma'\omega^{3/2}/\sqrt{2}c^2\} \\ q&=(B''\gamma'\sqrt{\omega}/2c\sqrt{r})\exp\{-i3\pi/4+r(1-i)\gamma'\omega^{3/2}/\sqrt{2}c^2\} \\ \theta'&=(B''/c\sqrt{r})\exp\{r(1-i)\gamma'\omega^{3/2}/\sqrt{2}c^2\}\end{aligned}\right\}. \qquad \ldots\ldots(65)$$

These are illustrated in fig. 4. The amplitude falls from a large value near the wall to a very small value near the axis. The thickness of the 'clinging' layer, measured to a point where the real part of the exponent is unity, is

$$\delta = \sqrt{2c^2/\gamma'\omega^{3/2}} = 1{\cdot}84 \times 10^8 f^{-3/2} \text{ cm for air.} \qquad \ldots\ldots(66)$$

The discriminant between the wide and very wide tube types of propagation is thus

$$\frac{r_w \omega^{3/2}\gamma'}{c^2} \propto \frac{r_w}{\delta} \propto \text{modulus of } Q_1 \propto \frac{\text{free gas absorption}}{\text{tube absorption}} \sim 10^{-8} r_w f^{3/2}. \qquad \ldots\ldots(67)$$

At a frequency of 100 c/s δ is over a mile in thickness. Thus to make measurements in the wide to very wide tube transition region, tubes of the order of a centimetre radius and frequencies of the order of 10 kc/s are suitable. To keep below the cut-off frequency of the higher-order modes and also inside the transition region it would be useful to work at low pressures.

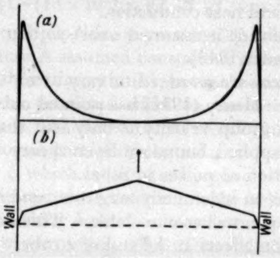


Fig. 4. Cross section profiles in a very wide tube, (*a*) amplitude of axial particle velocity or temperature fluctuation, (*b*) wave front or equiphase surface.

The wave front is inclined to the normal to the wall at an angle $(\gamma'/c)(\omega/2)^{1/2}$ which is very small, and it is convex towards $+x$ showing that the wave is divergent. The last term in m' is equal to the product of the 'attenuation' constant in the 'clinging' layer and the inclination of the wave front:

$$\gamma'^2\omega^2/2c^3 = \gamma'\omega^{3/2}/\sqrt{2c^2} \times (\gamma'/c)(\omega/2)^{1/2} \qquad \ldots\ldots(68)$$

so that this extra attenuation corresponds to an enhanced energy transfer to the walls.

Rayleigh (1896, p. 327, 1901) has shown that there is a similar type of propagation between parallel rigid conducting plane walls, and obtains a term equivalent to the last in (64), but leaves out the second.

(v) *Other Solutions of the General Equation* (33)

For very narrow tubes it is no longer permissible to neglect m^2 in comparison with h/ν. If we follow in other respects the narrow tube development we find

$$(r_w^2 c^2/8\gamma h^2)m^4 - (r_w^2 c^2/8\gamma\nu h)m^2 + 1 = 0. \qquad \ldots\ldots(69)$$

There are three limiting cases, one expressed by (36), and the others

$$m^2 = 2\omega(2\gamma)^{1/2}/cr_w \quad \ldots\ldots(70) \qquad\qquad m^2 = h/\nu. \quad \ldots\ldots(71)$$

Equation (36) is Rayleigh's narrow tube solution, which passes over into (70) as the radius approaches the mean free path λ'. The classical continuum theory is able to predict a correction at the radius (about 10^{-5} cm) where its foundations begin to break down because the numerical value of ν has been used, and this depends on λ'. There is a similar prediction when the wavelength approaches the mean free path. Using the resistance concept, and applying Knudsen's formula to (5), we find

$$m^2 = 3h(2\pi\gamma)^{1/2}/8cr_{\mathrm{w}} \qquad \ldots\ldots(72)$$

which is very similar to (70). The solution expressed by (71) suggests that viscosity waves propagating along the tube axis can co-exist with acoustic waves. There are also other solutions corresponding to temperature waves and to higher order modes in the tube.

§ 5. THE VALIDITY OF KIRCHHOFF'S FORMULAE

The assumptions which Kirchhoff made in deriving the wide tube formulae (1) and (2), in addition to those noted in table 1, are briefly stated here.

(*a*) Kirchhoff assumes a homogeneous medium, and only considers the extra effects due to viscosity and heat conduction.

(*b*) The sound amplitude is assumed small, so that there is no circulation or turbulence etc. (see Binder 1943).

(*c*) All the parameters are assumed to vary as $e^{i\omega t - mx}$, so that the formulae do not apply to pulses. Henry (1931) has pointed out that for a wave train the tube correction for the group velocity is only half that for the phase velocity. Boussinesq (1891) has applied boundary layer theory to give a comprehensive account of the propagation of pulses in tubes.

(*d*) The theory is for an indefinitely long tube, and it may take some distance for the velocity profile etc. to become stable. Thiesen (1907) has investigated the extra heat conduction effects at the end of a tube. Schweikert (1915, 1917) has considered the positions of the nodes in a Kundt's tube, and predicts in some circumstances an apparent velocity reduction due to absorption and proportional to $f^{-3/2}$. This agrees with the experiments of Schneebeli (1869) and Seebeck (1870), which have long been difficult to interpret.

(*e*) Circular symmetry is assumed, so that the wall surface is given by $r = r_{\mathrm{w}}$, the surface must also be perfectly smooth, impervious and rigid. Fissures parallel to the axis will increase the tube effect by increasing the ratio E/S (see eqn. (28)), and corrugations may cause an irregular gas motion at the wall. An imperfect wall surface is a frequent cause of experimental disagreement with Kirchhoff's formulae.

(*f*) The motion is assumed to vanish at the walls. The effect of non-rigid walls has been investigated by many workers (e.g. Korteweg 1878), and may be large at resonance or for a rubber-like tube. Henry (1931) considers the effect of a relative normal motion of tube and gas at the wall, such as might occur for a porous material etc. He concludes that this will lead to only a small velocity change, but may cause appreciable extra absorption. Henry also allows for molecular slip by modifying the last part of Kirchhoff's analysis, and his result together with that for the temperature discontinuity effect (paragraph (*g*)) may be written

$$\begin{aligned} c' &= c\left\{1 - \frac{\gamma'}{r_{\mathrm{w}}(2\omega)^{1/2}} + \left(\frac{\pi}{2P\rho}\right)^{1/2}\frac{\mu}{r_{\mathrm{w}}}\,\frac{2-f'}{f'} + \frac{(\gamma-1)\sqrt{\gamma k}}{r_{\mathrm{w}}\zeta}\right\} \\ &= 3{\cdot}434\times 10^4\left\{1 - \frac{0{\cdot}162}{r_{\mathrm{w}}\sqrt{f}} + \frac{6{\cdot}5\times 10^{-6}}{r_{\mathrm{w}}}\,\frac{2-f'}{f'} + \frac{4{\cdot}4\times 10^{-6}}{r_{\mathrm{w}}}\,\frac{2-g'}{g'}\right\}\ \text{cm sec}^{-1}\text{ for air} \end{aligned} \qquad \ldots\ldots(73)$$

where according to Knudsen the thermal permeability

$$\zeta \sim (k/1{\cdot}9\lambda')\{2g'/(2-g')\} \qquad \ldots\ldots(74)$$

and λ' is the gaseous mean free path, f' and g' the coefficients of slip and accommodation. The correction terms in the attenuation formula due to slip, temperature discontinuity, and a cross-product term are all found to contain $\omega\nu/c^2$, and are therefore negligible at normal frequencies.

The corresponding expressions for a narrow tube are given by eqn. (5), using a value of R modified for slip:

$$c' = cr_w\left(\frac{\pi f}{2\gamma\nu}\right)^{1/2}\left\{1+\left(\frac{2\pi}{P\rho}\right)^{1/2}\frac{\mu}{r_w}\,\frac{2-f'}{f'}\right\}$$

$$= 9{\cdot}346\times10^4\, r_w\sqrt{f}\{1+(13\times10^{-6}/r_w)(2-f')/f'\}\ \text{cm sec}^{-1} \text{ for air} \qquad \ldots\ldots(75)$$

$$m' = \frac{2(2\pi\gamma\nu f)^{1/2}}{cr_w}\left\{1-\left(\frac{2\pi}{P\rho}\right)^{1/2}\frac{\mu}{r_w}\,\frac{2-f'}{f'}\right\}$$

$$= 6{\cdot}701\times10^{-5}(\sqrt{f}/r_w)\{1-(13\times10^{-6}/r_w)(2-f')/f'\}\ \text{cm}^{-1} \text{ for air.} \qquad \ldots\ldots(76)$$

The correction term due to slip is twice as great in (75) as in (73).

(*g*) The gas temperature is assumed constant at the walls. Henry takes into account the finite temperature discontinuity, but makes some errors in an algebraic transformation. Thus the predicted effect on the attenuation is too large by several orders of magnitude, and the corrected velocity may be written as in (73). Zwikker (1941) also performs part of this calculation but it is not clear where he obtains the low value for ζ, which invalidates his conclusions.

There is also a small temperature variation in the wall at its surface. Since it has a finite thermal conductivity k_w and finite volume specific heat $\rho_w c_w$, rapidly damped temperature waves are propagated into the wall thickness. Henry (1931) and Nielsen (1949*a*) have shown independently that for both attenuation and velocity γ' is changed very slightly

$$\gamma' = \sqrt{\nu} + (\gamma-1)(\nu'/\gamma)^{1/2}\{1-(k\rho c_p/k_w\rho_w c_w)^{1/2}\}. \qquad \ldots\ldots(77)$$

(*h*) Kirchhoff assumes that the velocity of propagation of viscous effects $(2\nu\omega)^{1/2}$ or of temperature waves $(2\nu'\omega/\gamma)^{1/2}$ is much less than the velocity of sound c; i.e. $\omega\nu/c^2 \ll 1$ and $\omega\nu'/\gamma c^2 \ll 1$. This is again equivalent to the sound wavelength being much greater than the mean free path, which approximation breaks down at about 10^9 c/s for normal pressures.

(*i*) Kirchhoff's formulae only apply to waves of the principal mode. Rayleigh (1896, p. 161) has shown that a wave must eventually become plane if the frequency be lower than the lowest natural transverse frequency, so for a circular tube it is desirable that

$$\lambda > 3{\cdot}413 r_w. \qquad \ldots\ldots(78)$$

§ 6. APPLICATION OF THE THEORY TO LAWLEY'S RESULTS

The application of the theory to the recent work of Lawley (1952) is of interest, especially to the part with varying pressure. He measures the attenuation in oxygen at 120 kc/s, in a tube of diameter 1·5 mm, and at pressures from 5 to 130 cm Hg. That part of the absorption proportional to $P^{-1/2}$ (tube absorption, K_T) can be separated from that proportional to P^{-1} (free gas absorption, K_G) by plotting $m'P^{1/2}$ against $P^{-1/2}$.

Examination of eqns. (45) and (60) shows that although the tube is 'wide' all the time, the corrections arising because of the transition to both the narrow and very wide tube types are of a magnitude comparable with the free gas absorption.

The complete expression is

$$m' = \frac{\gamma'(\pi f)^{1/2}}{c r_{\mathrm{w}}} + \frac{\gamma''}{c r_{\mathrm{w}}{}^{2}} + \frac{2\pi}{c^3}\left\{\nu + \nu'' + \frac{\gamma - 1}{\gamma}\nu'\right\} f^2 + \frac{\pi^2 \gamma'^2 f^2}{c^3} \qquad \ldots\ldots(79)$$

or

$$m' P^{1/2} = K_{\mathrm{T}} + K_{\mathrm{G}} P^{-1/2}$$

$$= 1{\cdot}25 + (0{\cdot}068 + 0{\cdot}175 + 0{\cdot}101) P^{-1/2} \qquad \ldots\ldots(80)$$

for oxygen at 27°c, where P is in cm Hg.

It is now found that the theoretical value for K_{G} is *greater* than the experimental figure (table 2). Probably the main effect, as yet unconsidered, is the molecular absorption, in particular the manner in which it varies with pressure. The latter depends on the relaxation frequency at atmospheric pressure.

Table 2. Values of Constants in equation (80)

	K_{T}	K_{G}
Experimental (Lawley)	1·38	0·246
Theoretical	1·25	0·344

ACKNOWLEDGMENT

The author is indebted to Dr. R. W. B. Stephens for his interest in the work and for many helpful discussions.

REFERENCES

BALLANTINE, S., 1932, *J. Acoust. Soc. Amer.*, **3**, 319.

BINDER, R. C., 1943, *J. Acoust. Soc. Amer.*, **15**, 41.

BOGERT, B. P., 1950, *J. Acoust. Soc. Amer.*, **22**, 432.

BOUSSINESQ, 1891, *J. Phys.* (ser. 2), **10**, 301.

BRADFIELD, G., 1951, *Overdruk int het Colloquium over Ultrasonore Trillingen* (Brussels : Vlaam. Acad. Wet.), p. 199.

CARRIERE, Z., 1929, *J. Phys.* (ser. 6), **10**, 198.

CRANDALL, I. B., 1927, *Theory of Vibrating Systems and Sound* (New York : Van Nostrand).

CREMER, L., 1948, *Arch. elekt. Übertragung*, **2**, 136.

DANIELS, F. B., 1947, *J. Acoust. Soc. Amer.*, **19**, 569 ; 1950, *Ibid.*, **22**, 563.

GOLAY, M. J. E., 1947, *Rev. Sci. Instrum.*, **18**, 347.

GOLDSTEIN, S. (Ed.), 1938, *Modern Developments in Fluid Dynamics* (Oxford : University Press).

HELMHOLTZ, H., 1863, *Verhandlungen des Naturhistorisch-medicinischen Vereins zu Heidelberg*, **3**, 16.

HENRY, P. S. H., 1931, *Proc. Phys. Soc.*, **43**, 340.

IBERALL, A. S., 1950, *Bur. Stand. J. Res., Wash.*, **45**, 85.

KIRCHHOFF, G., 1868, *Ann. Phys., Lpz.*, **134**, 177.

KONSTANTINOV, B., 1939, *J. Tech. Phys., U.S.S.R.*, **9**, 226.

KORTEWEG, D. J., 1878, *Ann. Phys. Chem.*, **5**, 525.

KOSTEN, C. W., 1949, *Appl. Sci. Res.* B, **1**, 241.

KUNDT, A., 1868, *Ann. Phys., Lpz.*, **135**, 337.

LAMB, H., 1898, *Mem. Manchr. Lit. Phil. Soc.*, **42**, No. 9.

LAWLEY, L. E., 1952, *Proc. Phys. Soc.* B, **65**, 181.

NIELSEN, A. K., 1949 a, *Mikrofonmaalinger* (Copenhagen : Tekniske Højskole); 1949 b, *Trans. Danish Acad. Tech. Sci.*, No. **10**.

RAYLEIGH, Lord, 1896, *The Theory of Sound*, II, 2nd edn. (London : Macmillan); 1901, *Mag.*, **1**, 301.

RICHARDSON, E. G., 1928, *Proc. Phys. Soc.*, **40**, 206.

RICHARDSON, E. G., and TYLER, E., 1929, *Proc. Phys. Soc.*, **42**, 1.

SCHNEEBELI, 1869, *Ann. Phys., Lpz.*, **136,** 296.
SCHWEIKERT, G., 1915, *Ann. Phys., Lpz.*, **48,** 593 ; 1917, *Ibid.*, **52,** 333.
SEEBECK, A., 1870, *Ann. Phys., Lpz.*, **139,** 104.
SEXL, T., 1928, *Ann. Phys., Lpz*, **87,** 570 ; 1930, *Z. Phys.*, **61,** 349.
STOKES, G. G., 1851, *Trans. Camb. Phil. Soc.*, **9,** 8.
THIESEN, M., 1907, *Ann. Phys., Lpz.*, **24,** 401.
ZWIKKER, C., 1941, *Physica*, **8,** 1102.
ZWIKKER, C., and KOSTEN, C. W., 1949, *Sound Absorbing Materials* (Amsterdam : Elsevier).

RESEARCH NOTES

Note on the Grazing-Incidence Integral Ch(x, χ) for Monochromatic Absorption in an Exponential Atmosphere*

By SYDNEY CHAPMAN †
Queen's College, Oxford

MS. received 6*th May* 1953

This Note is a sequel to two papers (Chapman 1931 a, b, to be referred to as A and B) which dealt, *inter alia*, with the *absorption* of monochromatic solar radiation in an exponential atmosphere—that is, one in which the density ρ at height h is given, in terms of that at the ground, ρ_0, by

$$\rho = \rho_0 \exp(-h/H),$$

where H, there assumed independent of h, is the scale height (Chapman 1937).

In paper A the level layers of the atmosphere were treated as plane; in paper B I took account of the curvature of the layers (radius of curvature $a+h$, where a denotes the radius of the earth). I showed that in case A the rate of absorption I when the sun's zenith distance is χ is expressible as

$$I = I_0 \exp(1 - z - e^{-z} \sec \chi), \qquad \ldots\ldots(1)$$

where

$$z = (h - h_0)/H, \qquad \ldots\ldots(2)$$

and h_0 is the height of the level of maximum rate of absorption I_0 at vertical incidence ($\chi = 0$). In case B the equation (1) must be notified by the substitution, for $\sec \chi$, of a function which I denoted by $f(x, \chi)$, but which in recent years has been denoted by other workers by the symbol $\mathrm{Ch}(x, \chi)$, here also adopted. It is defined by

$$\mathrm{Ch}(x, \chi) = x e^{x} \sin \chi \int_0^{\chi} \exp(x - x \sin \chi / \sin \lambda) \operatorname{cosec}^2 \lambda \, d\lambda, \qquad \ldots\ldots(3)$$

and in its application to the present problem

$$x = R + z, \qquad \ldots\ldots(4)$$

where

$$R = (a + h_0)/H. \qquad \ldots\ldots(5)$$

After two decades radio scientists are finding useful applications of this function, and for this reason I present in this Note some considerations concerning it, which are among the results gained in efforts I have made from time to time since 1931 to extend and improve my original discussion of it‡.

---

* The work described in this Note was supported in part by the U.S.A. Signal Corps.

† Now at The Geophysical Institute, University of Alaska.

‡ Corrigenda may be noted as follows: to A, p. 32, table 1, the values of $z(\delta, \theta)$ need to be multiplied by 2·303, as they represent $\log \operatorname{cosec}(\theta+\delta)$ instead of $\ln \operatorname{cosec}(\theta+\delta)$; A, p. 40, in lines 10, 11, the numbers 0·8 and 0·9 should be 0·08 and 0·09; to B, p. 489, in the first line of eqn. (29) on the right, the sign $\mp$ should be replaced by $\pm$; p. 492, in eqn. (41) a term $x/\sin \chi$ should be added on the right; p. 493, table 4, the values of f should be amended as follows, with corresponding changes in $\ln f$: $\chi = 75°$, $R = 50$, $f = 3{\cdot}238$; $R = 100$, $f = 3{\cdot}499$; for $\chi = 80°$, $R = 50$, 100, 200, $f = 4{\cdot}169$, $4{\cdot}719$, $5{\cdot}109$.

These corrections were found necessary by Dr. W. C. Chiu in December 1950, while working with me at the California Institute of Technology, Pasadena, on a Signal Corps contract; Dr. Chiu also checked the approximate conformity of the values of $f(x, \chi)$ and $f(x, \pi - \chi)$ in B, table 4, with the formula (6) of this Note. The values of f in B, table 4, cannot all be relied on to four significant figures.

In paper B the values of $f(x, \chi)$ in table 4 for values of χ greater and less than 90° were calculated independently; but I should have noted that this was unnecessary, by virtue of eqn. (29). In the present notation, the addition of the two forms of (29), got by taking separately the signs + and − (where the equation gives ± or ∓), leads at once to the result (for $\chi<90°$)

$$\mathrm{Ch}(x, \pi-\chi)=\mathrm{Ch}(x, \chi)+2\mathrm{e}^{x(1-\sin \chi)}\mathrm{Ch}(x \sin \chi, 90°). \qquad \ldots\ldots(6)$$

In the last term, putting $y=x\sin\chi$, by B (16),

$$\mathrm{Ch}(y, 90°)=-y\mathrm{e}^{y}K_1(y),$$

where K_1 denotes the well known Bessel function (cf. Whittaker and Watson 1915, p. 377, ex. 40); for the values of y appropriate to the ionosphere, $\mathrm{Ch}(y, 90°)$ is more easily calculable from the asymptotic formula

$$(\tfrac{1}{2}\pi y)^{1/2}\{1+\sum_{n=1} a_n y^{-n}\},$$

where $$a_n=[\{1\cdot 3\ldots(2n-3)\}^2(4n^2-1)]/(n!\,2^{3n}),$$

so that $$a_1=3/8, \qquad a_2=-15/128, \qquad a_3=105/1024 \quad \ldots.$$

Thus if $\mathrm{Ch}(x, \chi)$ has been calculated for a value of χ less than 90°, and for $\chi=90°$ for a suitable range of values of x, (6) enables the value of $\mathrm{Ch}(x, \pi-\chi)$ to be readily deduced therefrom.

As χ increases towards and beyond 90°, $\mathrm{Ch}(x, \chi)$ rapidly increases; it increases also, for any value of χ, with x. The values of $\mathrm{Ch}(x, \chi)$ in B, table 4, are too widely spaced for convenient interpolation, and for this purpose an alternative mode of tabulation of the function seems desirable.

One such mode may be afforded by introducing the new variable $\Delta\chi$ thus defined: $\mathrm{Ch}(x, \chi)=\sec(\chi-\Delta\chi)$.

Clearly $\Delta\chi$ is a function of both χ and x. From B, table 4, it is easy to calculate $\Delta\chi$ from the values of $\mathrm{Ch}(x, \chi)$ there given. This has been done for me by Dr. W. C. Chiu, with the results given in the table of this Note. From this table graphs of $\Delta\chi$ as a function of $\log(R/50)$ for different values of χ (e.g. 30°, 45°, 60°, 75°, 80°, 83°, 85°) can be drawn, which will be found useful for interpolating to give $\mathrm{Ch}(x, \chi)$ for intermediate values of x and χ. To find $\mathrm{Ch}(x, \chi)$ for values of χ greater than 90° it is preferable to determine $\mathrm{Ch}(x, \pi-\chi)$ and $\mathrm{Ch}(y, 90°)$ by interpolation, and then to use (6).

Table of Approximate Values of $\Delta\chi$

	$\chi=30°$	45°	60°	75°	80°	83°	85°	87°	90°
$R=50$	35′	1° 3′	1° 36′	3° 1′	3° 53′	4° 25′	4° 53′	5° 26′	6° 26′
100	20′	33′	55′	1° 36′	2° 14′	2° 45′	3° 8′	3° 37′	4° 34′
200	11′	17′	28′	55′	1° 17′	1° 38′	1° 56′	2° 21′	3° 14′
300	7′	11′	19′	39′	55′	1° 11′	1° 26′	1° 48′	2° 38′
400	5′	8′	15′	30′	43′		1° 9′	1° 28′	2° 17′
500	4′	6′	12′	24′	35′		58′	1° 15′	2° 3′
600	4′	5′	10′	20′	29′		50′	1° 6′	1° 51′
650	4′	5′	9′	19′	27′		47′	1° 3′	1° 48′
700	4′	5′	8′	17′	25′		44′	59′	1° 44′
800	4′	5′	7′	15′	22′		39′	54′	1° 37′

Dr. M. V. Wilkes informs me that he is preparing, by numerical quadrature, an extended table of the Ch function for values of χ at 1° intervals.

REFERENCES

CHAPMAN, S., 1931 a, *Proc. Phys. Soc.*, **43**, 26; 1931 b, *Ibid.*, **43**, 483; 1937, *Reports on Progress in Physics*, **3**, 42 (London : Physical Society).

WHITTAKER, E. T., and WATSON, G. N., 1915, *Modern Analysis*, 2nd Edn. (Cambridge : University Press).

Ferromagnetic Domain Processes in Single Crystal Disc Specimens of Silicon Iron

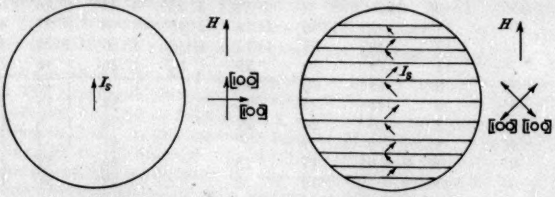
BY D. H. MARTIN
University of Nottingham

Communicated by L. F. Bates; MS. received 20*th March* 1953

IN a recent paper (Bates and Martin 1953) a domain process resulting in the demagnetization of a single crystal of silicon iron, initially saturated in an easy direction, was described. Imperfections (pits) in the specimen's surfaces were there observed to play an important role, and the importance of surface imperfections in another domain process is described below.

A disc specimen (diameter 7·8 mm, thickness 0·36 mm) was cut from a single crystal of 3% silicon iron, the surfaces being parallel to the (100) plane. It was electrolytically polished in the usual way for the study of domain structures using the powder deposit technique (cf. Bates and Mee 1952).

This specimen was placed in a uniform magnetic field which was at first parallel to a [100] direction in the plane of the specimen. It was then rotated in the field about a perpendicular axis at right angles to the field until the latter was along a [110] direction. The field was sufficient to saturate the specimen while it was in the initial position, but not when it was rotated into the final position. The main features of the domain structures for these two conditions are illustrated diagrammatically in figs. 1 (*a*) and (*b*), which are self-explanatory.



(*a*) (*b*)

Fig. 1. Domain structures in disc specimen.

The domain process bringing about this change in structure as the specimen was rotated was recorded. A point of particular interest was the origin of those domains whose direction of magnetization had no equivalent in the initial condition of fig. 1 (*a*). The field had first been increased until saturation in the [100] direction (an 'easy direction') had been attained, a condition satisfactorily indicated by the disappearance of all powder patterns from the surface of the disc. On rotating the specimen through a small angle, small dagger-shaped closure structures appeared at surface imperfections. The larger the imperfection the smaller in general was the rotation required for the formation of such a structure. The sudden appearance of many powder patterns with daggers already of some size suggests that the latter are created by an irreversible process.

Further rotation of the specimen resulted in appreciable growth of these dagger structures until, when the applied field lay in a [110] direction, many had spread across the whole width of the surface. This process is illustrated diagrammatically in fig. 2. Figure 2 (*a*) shows an imperfection with no dagger closure structure; the specimen is saturated and the applied field and [100] directions are parallel. Figure 2 (*b*) shows the appearance of a dagger structure as a result of rotation through a few degrees. Figures 2 (*c*) and (*d*) show the progressive growth of this structure with further rotation. Dagger structures from other imperfections and from the edges of the specimen (see below) merge with or grow alongside one another so that when eventually the field lies along a [110] direction the domain configuration of the type shown in fig. 1 (*b*) is attained, as illustrated in figs. 2 (*e*) and (*f*).

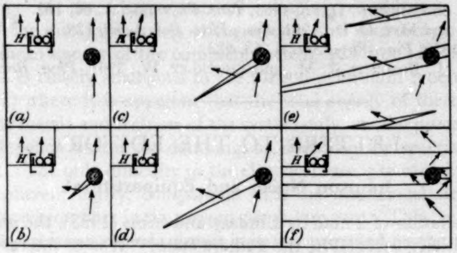


Fig. 2. Growth of new domains.

Some photographs of the growth of a dagger are shown in figs. 3 (*a*), (*b*) and (*c*) (Plate), which correspond approximately to the conditions depicted in figs. 2 (*b*), (*c*) and (*d*) respectively. There is a tendency for powder deposits to form on one side of a dagger only (see Mee 1950), and thus the daggers usually appear in the photographs merely as a single line. An imperfection and the base of a dagger structure are shown under higher magnification in fig. 3 (*e*), where the striations of the deposit (cf. Williams, Bozorth and Shockley 1949) show that the directions of magnetization inside and outside the dagger are *not* mutually perpendicular. This is to be expected when a field is acting on a specimen.

Many similar dagger structures originated at regions on the edges of the specimen. These behaved in the same way as those appearing at surface imperfections, and are believed to be due to imperfections of a similar nature which must be numerous close to the edges of a specimen. A photograph of the patterns at one of these is shown in fig. 3 (*d*), where a number of daggers are seen leaving the edge.

Since powder patterns directly reveal only the domain structures of the surface layers it is to be remembered that the technique by itself cannot demonstrate *conclusively* that the dagger structures finally spread throughout the thickness of the specimen. The patterns obtained, however, make this appear highly probable. It is also not at present possible to determine whether similar processes to those observed on the surface also originate in the interior of the specimen. It is hoped that other experiments will provide conclusive information to settle these problems.

ACKNOWLEDGMENTS

This work was carried out in the Physics Department of the University of Nottingham and I wish to record my thanks to Professor L. F. Bates for his constant interest, and to the Electrical and Allied Industries Research Association for financial support which made it possible and for permission to publish.

REFERENCES

BATES, L. F., and MARTIN, D. H., 1953, *Proc. Phys. Soc.* A, **66,** 162.
BATES, L. F., and MEE, C. D., 1952, *Proc. Phys. Soc.* A, **65,** 129.
MEE, C. D., 1950, *Proc. Phys. Soc.* A, **63,** 922.
WILLIAMS, H. J., BOZORTH, R. M., and SHOCKLEY, W., 1949, *Phys. Rev.*, **75,** 155.

LETTERS TO THE EDITOR

Johnson Noise and Equipartition

At the conclusion of a note by Lindsay and Sims (1953), the general content of which is entirely acceptable, the authors state: "It is of interest to note that no reactive terms appear in any of these expressions [for the Nyquist thermal-noise relation]". It is true that reactive terms do not appear explicitly, but the first expression cited is for the total fluctuation (all frequencies) in the form

$$\overline{V^2}=\left(\frac{2}{\pi}\right)\boldsymbol{k}T\int\mathscr{R}\left(\frac{1}{Y(\omega)}\right)d\omega.$$

For comparison with the original Nyquist relation $\overline{V^2}_{df}=4R\boldsymbol{k}T\,df$ this can be written as

$$\overline{V^2}=(2\boldsymbol{k}T/\pi)\int\mathscr{R}[Z(\omega)]\,d\omega$$

and Moullin and Ellis (1934) have shown that in the simple example of an L, C, R combination the integral of $\mathscr{R}[Z(\omega)]$ over all frequencies depends only on the shunt capacitance and not on the resistance. It follows that $\overline{V^2}=\boldsymbol{k}T/C$ which may be interpreted as $\frac{1}{2}C\overline{V^2}=\frac{1}{2}\boldsymbol{k}T$, i.e. the complete circuit behaves as an equilibrium thermodynamic system with one degree of freedom in relation to equipartition.

More recently the network-analysis methods of Bode (1945) have made electrical engineers aware that one can generalize this result for passive networks of any form provided only that they consist of lumped elements and their frequency characteristics do not have poles on the axis of real frequencies, i.e. that they include some element of dissipation. (Since only two-terminal impedances are in question, the possibility of non-minimum-phase lumped networks does not arise.) Any two-terminal circuit in which the shortest geometric path between terminals is occupied by a dielectric will reduce to a shunt capacitance C' as $\omega \to \infty$ and in that case Bode's result I(a) on p. 301 indicates that

$$\int_0^\infty \mathscr{R}[Z(\omega)]\,d\omega = \pi/2C'. \qquad \ldots\ldots(1)$$

Similarly, if the shortest path between terminals is occupied by a conductor the circuit admittance will tend to an inductive susceptance at infinite frequency and the corresponding admittance analysis will lead to

$$\int_0^\infty \mathscr{R}[Y(\omega)]\,d\omega = \pi/2L'. \qquad \ldots\ldots(2)$$

Thus the infinite integrals of the formulae of the Nyquist type which explicitly appear independent of reactive components are implicitly expressible in terms of *reactive components alone*, though this is on condition that there shall be some finite amount of dissipation in the circuit. This is the analytical expression of the physical equivalence between *micro*scopic equipartition, the thermal energy of the charge carriers reaching equilibrium through the mechanism of the dissipative parts of the circuit, and *macro*scopic equipartition in which the mean energy of the complete system, $\frac{1}{2}C\overline{V^2}$ or $\frac{1}{2}L\overline{I^2}$ is found equal to $\frac{1}{2}\boldsymbol{k}T$. This is closely analogous to the torsion-pendulum type of experiment (Kappler 1931) where it is apparent that the total energy of thermal oscillation depends on the inertia and stiffness of the system only, yet the process of reaching equilibrium depends on the air damping, as does also the spectral distribution of the energy. The only difficulty in the electrical case is to visualize the electric circuit as a coherent entity, comparable with the mechanical structure of the mirror and suspension fibre in the torsion-pendulum case. This 'materialization' of the otherwise abstract electric circuit is in fact provided by the mutual energy of the electrons (or other charge carriers) which embody the electrical properties of the system of material conductors. The electrostatic energy $\frac{1}{2}CV^2$ represents the sum of the mutual potential energies between the individual charged particles which together constitute a conductive circuit. Similarly it has been shown by Brillouin (1934) that $\frac{1}{2}LI^2$ represents the sum of all the mutual magnetic energies (comparable with kinetic energies) of the moving charge carriers. In this way the fact that the 'circuit energy' is physically derived from the individual particle energies both gives meaning to the idea of the 'circuit' as an entity in equipartition theory and provides the physical parallel to the analytic relation between resistive and reactive parts which is revealed by Bode's theorems.

A detailed discussion of this topic, on lines similar to the above, is included in a paper by Fürth (1948).

Electrical Engineering Department,
University of Birmingham.
15th May 1953.

D. A. Bell.

Bode, H. W., 1945, *Network Analysis and Feedback Amplifier Design* (New York: Van Nostrand).
Brillouin, L., 1934, *Helv. Phys. Acta* (*Supplement*), **7**, 47.
Fürth, R., 1948, *Proc. Roy. Soc.* A, **192**, 593.
Kappler, E., 1931, *Ann. Phys., Lpz.*, **11**, 233.
Lindsay, P. A., and Sims, G. D., 1953, *Proc. Phys. Soc.* B, **66**, 423.
Moullin, E. B., and Ellis, H. D. M., 1934, *J. Instn. Elect. Engrs.*, **74**, 323.

The Influence of Cathode Material on the Electric Strength of Potassium Bromide Crystals

The safe working stress of electrical insulation is not, in practice, generally determined by the intrinsic strength of the material. But the problem of intrinsic breakdown exists and for many years it has been the subject of experimental and theoretical investigation. Amongst materials studied the alkali halides are prominent and theoretical values of their electric strength have been available for some time. However, whilst various investigators have carried out experimental investigations with these materials it cannot yet be said that there exist data of sufficient reliability. This is due possibly to the existence of a number of secondary effects that are not yet fully understood and the experiments of Alger and von Hippel (1949) suggest that one such effect is the influence of the cathode material. It is with this possible effect that the present letter is concerned.

Alger and von Hippel investigated at room temperature the influence of the electrodes on the intrinsic strength of potassium bromide crystals. They found that the average electric strength was increased by about 60% when electrodes consisting of a saturated solution of potassium bromide in water were used instead of mercury or gold electrodes. Alger and von Hippel interpreted this result by assuming that, with metal electrodes, space charge distortion of the electric field in the crystal was caused by field emission of electrons from the cathode. This effect is not possible with a cathode consisting of an ionic solution and hence higher values of electric strength were obtained.

We have carried out similar experiments with crystals of potassium bromide. Plane-recessed specimens were used similar to those of Alger and von Hippel and the cathode was either of graphite, gold or an ionic solution of potassium bromide. The method of preparing specimens and other experimental techniques have been described previously (Calderwood and Cooper 1953) and direct current was used to break down most of the specimens.

The average electric strength of 17 specimens with gold cathodes was 0·79 MV cm^{-1}; the spread in electric strengths extended from 0·56 MV cm^{-1} to 1·03 MV cm^{-1}. The average value obtained for a group of seven specimens with graphite cathodes was 0·72 MV cm^{-1} and it lay between a minimum value of 0·60 MV cm^{-1} and a maximum value of 0·85 MV cm^{-1}. When KBr solution was used for the cathodes of 15 specimens, the average electric strength was 0·73 MV cm^{-1}, the maximum value was 0·94 MV cm^{-1} and the minimum value was 0·52 MV cm^{-1}. Seven of the specimens of this last group were broken down with $\frac{1}{50}\,\mu$sec standard impulses but no difference could be detected between the impulse and the d.c. behaviour.

All of the specimens used in the above experiments were annealed and relatively strain free, as observed in a polarizing microscope, before the voltage was applied.

Application of Student's 't' test to the results of the experiments showed that no significance could be attached to the small differences between the above average values. This result contrasts with that of Alger and von Hippel. The result is, however, similar to that obtained recently for KCl crystals (Calderwood, Cooper and Wallace 1953).

A possible explanation of the difference between the present result and that of Alger and von Hippel is given below. Alger and von Hippel apparently did not treat identically their various groups of specimens. Those with metal electrodes were heat treated at 375°c but the ones with potassium bromide solution as electrodes were not heat treated. Some evidence exists for believing that the electric strength of alkali halide crystals may be reduced by previous heat treatment (Calderwood, Cooper and Wallace 1953), and it is possible that the results of Alger and von Hippel was primarily due to differences in the treatment before breakdown of the different groups of crystals and not to differences in the cathode material.

Electrical Engineering Department,
University of Liverpool.
9th June 1953.

R. Cooper.
D. T. Grossart.

Alger, R. S., and von Hippel, A., 1949, *Phys. Rev.*, **76**, 127.
Calderwood, J. H., and Cooper, R., 1953, *Proc. Phys. Soc.* B, **66**, 73.
Calderwood, J. H., Cooper, R., and Wallace, A. A., 1953, Symposium on Insulating Materials, *Proc. Instn. Elect. Engrs.*, **100**, Pt. IIA, No. 3, in the press.

The Surface Deformation caused on Iron Crystals by Unidirectional Abrasion

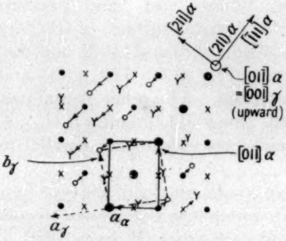
We have investigated by electron diffraction the surface deformation caused by very light unidirectional abrasion on electropolished, roughly (107) and (110), surfaces of 2 mm thick single-crystal sheets of iron of low carbon content. A single 10-inch stroke was made at about 3 in. $\sec^{-1}$ on 0000 emery paper, lubricated with benzene, under light hand pressure amounting to about 50 g cm^{-2}. The nearly (107) crystal surface, of ($h0l$) type and inclined at $8\frac{1}{2}$° to (001), was abraded along $[\bar{1}10]$, $[\bar{2}10]$ and [100], and the observed large rotations of the fragments were found to be about [110], [120] and [010] axes respectively, with little spread from the mean. The other crystal used had a practically (110) surface and was abraded along $[\bar{1}1\bar{2}]$ and $[\bar{1}11]$. This again caused large rotations about well-defined axes which were parallel to $[1\bar{1}\bar{1}]$ and an axis near $[\bar{3}3\bar{4}]$ respectively.

In addition to the above results mostly of the type found previously (cf. Evans, Layton and Wilman 1951), which may correspond to coherent lamellar fragments in rotational slip relation to each other (Wilman 1950, 1951), the following new effects of abrasion were brought to light. The first of these was observed in the case of abrasion of the (107) surface along $[\bar{1}00]$, which gave abnormally obscure disorientation until, after etching down to a level where the disorientation was reduced to only a few degrees rotation, the diffraction patterns indicated a well-defined rotation. This showed that some parts of the lattice had rotated about an axis estimated to be near [451], while other parts had rotated about the axis $[4\bar{5}1]$ symmetrically disposed relative to the (010) plane which was parallel to the abrasion direction and normal to the surface. This rotation about [451] appears to correspond to flexural translational slip on $(\bar{2}13)$ along the usual slip direction $[1\bar{1}1]$, since [451] is perpendicular to $[1\bar{1}1]$ and lies in a $(\bar{2}13)$ plane.

The second new and particularly striking observation is that some of the body-centred-cubic α-iron ($a = 2{\cdot}86$ Å) was transformed into face-centred-cubic γ-iron ($a = 3{\cdot}60 \pm 0{\cdot}01$ Å relative to α-iron, $a = 2{\cdot}86$ Å) in the surface region, as a result of the abrasion. This was shown clearly in most cases by the γ-iron ring pattern obtained after the surface had been etched for only about 45 seconds in a 1% picric acid solution in alcohol. After 3 minutes etching the surface usually yielded only slightly arced rings corresponding to normal α-iron with some traces of γ-Fe_2O_3; and after roughly 4 to 6 minutes total etching, the α-iron rings had usually broken up into clearly separated arcs corresponding to a range of rotation from the initial crystal orientation, about a well-defined axis.

The presence of the γ-iron indicates (i) that even in such light abrasion the metal surface had reached a temperature of about 900°C or more, at least locally, as indeed is clearly possible from the results of Bowden and his collaborators (see for example Bowden and Tabor 1950) although the possibility of formation of γ-iron does not seem to have been suggested; (ii) that the cooling of these regions was sufficiently rapid to allow this γ-iron to retain its structure after cessation of the abrasive action. This observation also emphasizes that untreated abraded metal surfaces, especially iron, are likely to be unsuitable for use as the starting point of, for example, studies of corrosion, since the state of the surface is not closely reproducible.

Still further new results shown by the above experiments concern the orientation relationship between the γ-iron crystals and the α-iron crystal from which they were formed. From electron-diffraction patterns of sharp spots obtained after etching the abraded (110) iron surface below the rotationally disorientated region, it was established to a high accuracy to be of the type (001) γ parallel to (01$\bar{1}$) α, and in this common plane [$\bar{1}$10] γ parallel to [$\bar{1}$11] α. This orientation (see figure) differs from the well-known conclusion of Kurdjumow and Sachs (1930) and Sachs (1932) in the transformation of a single crystal of face-centred-cubic austenite to the body-centred-tetragonal (nearly cubic) martensite as a result of cooling, and also from that observed by Nishiyama (1934–35) in the martensitic transformation of 30:70% Ni–Fe crystals during cooling. Further spots and lines in the diffraction patterns show that the γ-iron was much twinned in thin lamellae parallel to {111} γ.



Projection of the α-iron crystal lattice on (01$\bar{1}$)α (● at heights ±0, 1, 2, 3, . . . $\times \sqrt{2}\, a_\alpha$, and × at heights $\pm\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots \times \sqrt{2}\, a_\alpha$); and the transformation to one of the γ iron lattices (○ and Y at heights ±0, 1, 2, 3, . . . and $\pm\frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$ respectively, $\times c_\gamma$).

The mechanism of such transformations is as yet incompletely understood. In the present case, however, there is a natural and simple relationship between the two structures, in that a $(01\bar{1})\,\alpha$ plane contains a centred $\sqrt{2}$-rectangular atomic array and becomes after transformation a centred square array in a $(001)\,\gamma$ plane, as in the figure. The transformation can be formally represented as equivalent to a homogeneous shear on $(211)\,\alpha$ planes along $[\bar{1}11]\,\alpha$, combined with a homogeneous 6·0% extension along $[211]\,\alpha$ as in the figure, and a homogeneous contraction of 13·4% along $[01\bar{1}]\,\alpha$, i.e. $[001]\,\gamma$, to convert the face-centred-tetragonal lattice ($c/a=2/\sqrt{3}=1{\cdot}155$) into the face-centred-cubic γ lattice. Although the atomic movements must actually be less linear than in this idealized representation, it is to be noted that the net motion is along the $\langle 111\rangle\,\alpha$, i.e. $\langle 110\rangle\,\gamma$, rows as in normal translational slip in the body-centred-cubic α and face-centred-cubic γ lattices.

At this stage of etching, patterns were also obtained which showed only α-iron diffractions consisting, however, of groups of four sharp 'irrational' spots which indicated that the crystal surface had the form of smooth $\{110\}$ facets. Since such patterns were never observed from unabraded surfaces after etching or electropolishing, it is concluded that they represent the interfaces between the γ-iron and the unchanged α-substrate.

It is hoped to publish a detailed account of these results in the near future.

Laboratory of Applied Physical Chemistry,
Imperial College, London.
30th April 1953.

R. P. Agarwala.
H. Wilman.

Bowden, F. P., and Tabor, D., 1950, *The Friction and Lubrication of Solids* (Oxford: Clarendon Press).
Evans, D. M., Layton, D. N., and Wilman, H., 1951, *Proc. Roy. Soc.* A, **205,** 17.
Kurdjumow, G., and Sachs, G., 1930, *Z. Phys.*, **64,** 325.
Nishiyama, Z., 1934-5, *Sci. Rep. Tohoku Univ.*, **23,** 638.
Sachs, G., 1932, *Z. Metallkde*, **24,** 241.
Wilman, H., 1950, *Nature, Lond.*, **165,** 321; 1951, *Proc. Phys. Soc.* A, **64,** 329.

BOOK NOTICES

Radio-Frequency Power Measurements, by Roald A. Schrack. National Bureau of Standards Circular 536. Pp. iii + 16. (Washington, D.C.: U.S. Department of Commerce, 1953.) 15 cents.

Le Point de Vue du Physicien dans le Problème du Lissage des Courbes Expérimentales, by Pierre Vernotte. Pp. ii + 23. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air (No. N.T. 46), 1953.) 400 fr.

Sur L'Entretien des Oscillations des Eaux Portuaires sous L'Action de la Haute Mer, by John S. McNown. Pp. 44. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, 1953.) 600 fr.

Les Autoplissements et les Équations Intrinsèques des Fluides Visqueux, by François-Joseph Bourrières. Pp. 95. (Paris: Publications Scientifiques et Techniques du Ministère de l'Air, 1953.) 1200 fr.

CONTENTS OF SECTION A

PAGE

Dr. J. M. DANIELS. The Effect of Interactions in a Paramagnetic on the Entropy and Susceptibility 673

Dr. D. L. LIVESEY and Dr. C. L. SMITH. Photodisintegration Processes in Light Even–Even Nuclei yielding Alpha-Particles 689

Mr. D. J. LITTLER and Mr. E. E. LOCKETT. The Pile Neutron Absorption Cross Sections of Bismuth 700

Mr. E. R. WOODCOCK. A Method of Calculating Critical Size in Multi-Group Neutron Transport Theory for Some Simple Systems 705

Dr. R. H. DALITZ. The Decay of the τ-Meson 710

Dr. R. MCWEENY. The Diamagnetic Anisotropy of Large Aromatic Systems—V: Interpretation of the Results 714

Mr. J. M. DICKSON and Mr. D. C. SALTER. A Search for Polarization of High Energy Neutrons 721

Dr. R. J. BLIN-STOYLE. Matrix Elements in Radiative Transitions . . . 729

Dr. M. E. PILLOW. Intensity Distribution among Bands of the Herzberg System of O_2 733

Dr. M. E. PILLOW. Band Intensities in the CN Violet System 737

Mr. P. SWAN. The Elastic Scattering of Neutrons by Tritons and of Protons by ^{3}He 740

Mr. P. C. BANBURY, Dr. H. K. HENISCH and Mr. A. MANY. On the Theory of the Isothermal Hall Effect in Semiconductors 753

Dr. H. G. HOWELL. The Visible Emission Spectrum of Cl_2^+ 759

Research Notes:

Dr. D. M. S. BAGGULEY. Ferromagnetic Resonance in Colloidal Suspensions 765

Mr. W. T. LINK and Dr. D. WALKER. Light Output of Potassium Iodide Crystals under Bombardment by Heavy Charged Particles 767

Letters to the Editor:

Mr. H. C. ROWLINSON and Dr. R. F. BARROW. The Absorption Spectrum of Aluminium Monofluoride in the Schumann Region 771

Dr. W. M. LOMER and Mr. K. W. MORTON. The Electrostatic Energy of Boron Nitride 772

Mr. A. L. MATHUR and Dr. K. M. GATHA. The Born–Yang Model for High Energy Electron Scattering 773

Reviews of Books 775

Contents of Section B 776

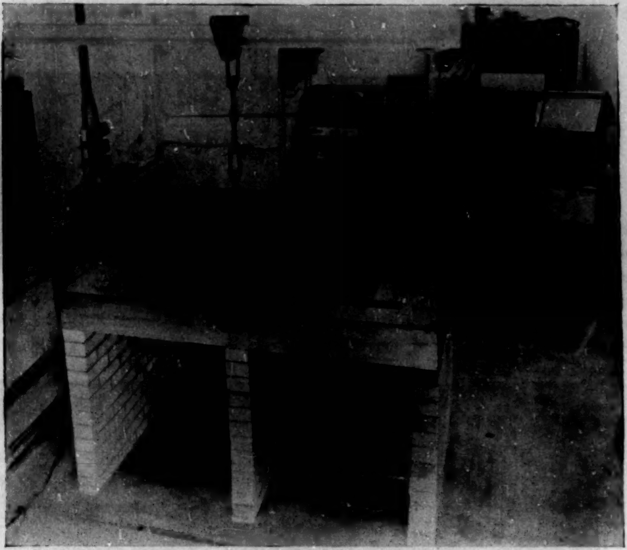


PLATE I. General view of microtron.

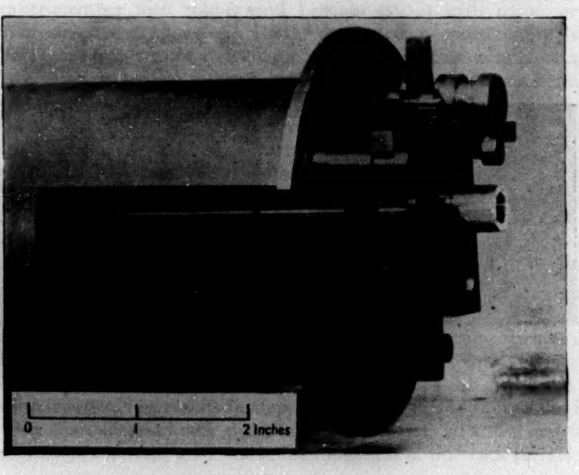


PLATE II. Extractor in position.

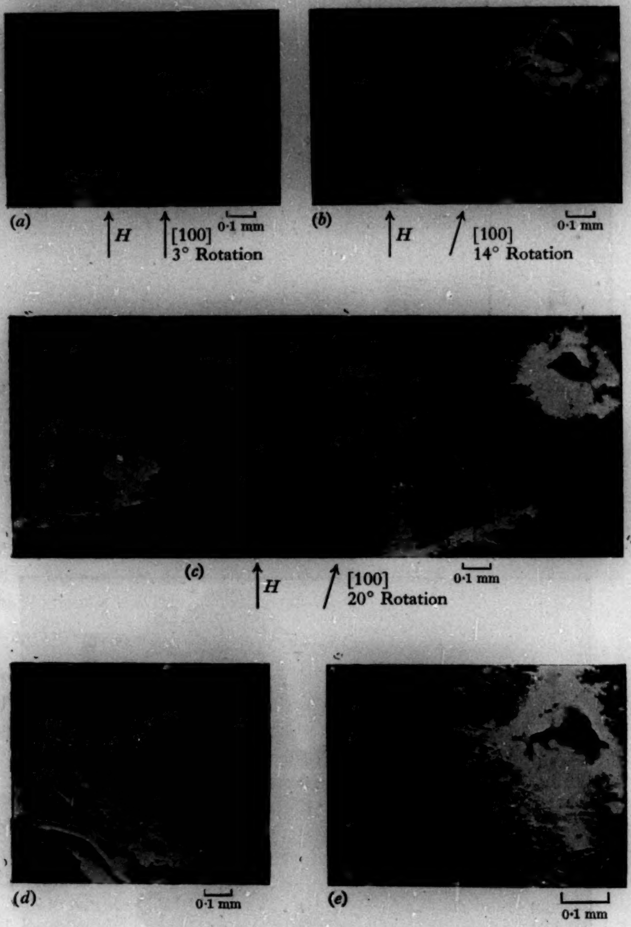

(a)
H
[100]
3° Rotation
0·1 mm
(b)
H
[100]
14° Rotation
0·1 mm
(c)
H
[100]
20° Rotation
0·1 mm
(d)
0·1 mm
(e)
0·1 mm

Fig. 3.

The Influence of Load and Surface Roughness on the Friction of Rubber-Like Materials

By D. F. DENNY
The British Hydromechanics Research Association, Harlow, Essex

MS. received 7th May 1953

Abstract. Experiments with rubber and gelatine under lubricated conditions confirmed that with rubber-like materials the load dependence of friction is consistent with elastic deformation of surface asperities and partial adhesion over the true area of contact. However, the $F \propto W^{2/3}$ relation found by previous investigators obtains only over a limited load range. At very low loads the friction is very nearly proportional to load, whilst at very high loads contact between specimen and track is complete, and the friction remains sensibly constant.

When rubber slides on rough tracks there is a large additional friction force unrelated to true contact area.

§ 1. INTRODUCTION

MANY investigators have reported that the friction of highly elastic materials does not increase in proportion to applied load. Of those who studied the subject in detail, Thirion (1946) observed that the friction of rubber appeared to obey the law $1/\mu = a + bp$, where a and b are constants and p is the apparent normal pressure. Subsequently Schallamach (1952) experimenting with rubber and Lincoln (1952) with nylon provided evidence in support of Bowden's (1951) suggested proportionality between friction force and true area of contact, for it was shown that over a certain load range the friction, as well as the area of contact between an elastic hemisphere and a rigid plane surface, varied as $W^{2/3}$.

However, it is clear that this latter relation can apply only over a limited load range. For instance, with rubber the deformation of the surface asperities will not continue indefinitely, and the area of contact must eventually reach a limiting value,* whilst with plastics the elastic limit will ultimately be exceeded and the area of contact thus be directly proportional to applied load.

At very low loads also it is unlikely that the friction coefficient will tend to the infinite value implied by the relation, and indeed the results of Thirion and Schallamach show that the friction does in fact fall below the value required by this relation.

The friction experiments described in this paper, covering various rubber compounds and gelatine materials, suggest that Thirion's law applies over a ten-thousand-fold load range and that it is also compatible with the elastic deformation theory. The results further indicate that in addition to the friction force arising from adhesion over the area of real contact, an additional force of comparable magnitude may arise due to the roughness of the track surface.

---

* This is implied in Thirion's formula.

§ 2. EXPERIMENTAL

The friction of nine rubber compounds and two gelatine-based solids (listed in table 1) was measured against various lubricated track surfaces. The quantity E_0, being the compression modulus (applied stress/linear strain) for strains below 10% was found to be convenient for the purpose of comparing the stiffness of one material with another.

Table 1. Details of Rubber and Gelatine Compounds

Material	(1)	(2)	(3)	(4)
Hycar	90	310	140	2·5
Hycar	75	126	126	2·4
Hycar	60	49	230	2·0
Neoprene	90	210	120	2·5
Reclaim (mixed)	80	98	74	2·8
Silicone	80	98	44	2·8
Silicone	60	45	35	1·7
Natural	60	42	98	2·8
Natural	40	14	250	2·5
Gelatine E	—	3·5	2·0	7·6
Gelatine B	—	0·7	0·5	6·2

(1) Nominal hardness (BS degrees) ; (2) compression modulus E_0 (kg cm^{-2}) ; (3) ultimate tensile strength (kg cm^{-2}) ; (4) coefficient A in eqn. (1).

Weight-loaded specimens were towed across the track at constant speed and the towing force was measured by a three-wire linkage, as shown in fig. 1, the wire being attached to the loading plate in the plane of the sliding interface, fig. 2.

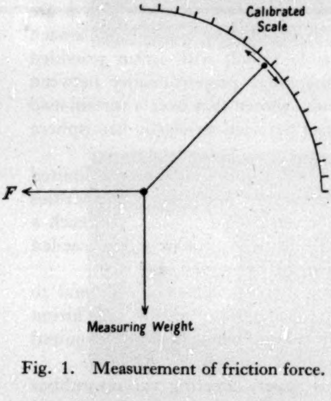


Fig. 1. Measurement of friction force.

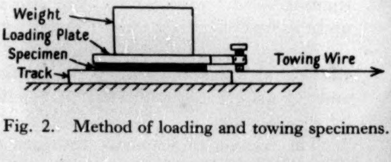


Fig. 2. Method of loading and towing specimens.

The specimens, cut to rectangular shape from sheets that had been moulded against polished steel, varied in size from 0·5 cm thick and 100 cm² area to 0·1 cm thick and 0·08 cm² area and were bonded to the loading plate. Weights up to 50 kg provided the required normal pressure. At very high pressures, when extremely thin specimens would have been necessary to avoid the increase of apparent surface area accompanying compressive strain, it was found more convenient to use cylindrical specimens enclosed in recesses in the loading plate, so that only about 0·01 cm protruded. Measurements were made at a speed of 0·01 cm sec^{-1}, below which velocity dependence of friction was negligible. Both specimen and track were washed in acetone before application of the lubricants and under these conditions measurements were repeatable within ±10%.

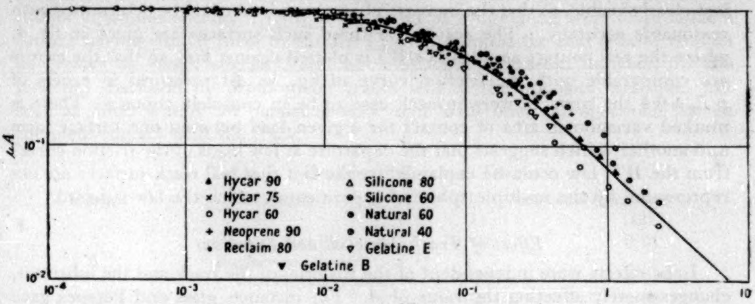


Fig. 3. Relation between coefficient of friction and apparent normal pressure. Polished steel track ; olive oil lubricant ; specimen materials as listed in table 1. Equation of full-line $1/\mu A=(1+15\ p/E_0)$.

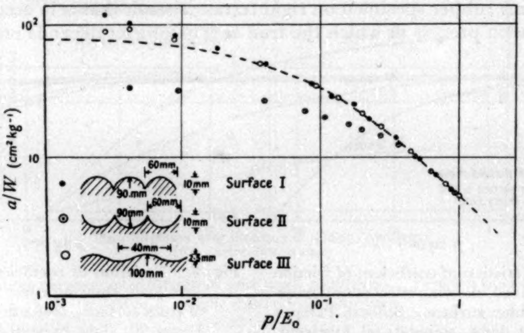


Fig. 4. Relation between area of real contact and apparent normal pressure for gelatine ($E_0=0{\cdot}2$ kg cm^{-2}) on wavy surfaces. Equation of dotted line $W/a \propto (1+15\ p/E_0)$.

Load Effect

Experiments indicated that the coefficient of friction depended only on the apparent normal pressure p (i.e. load W/area of specimen) and not on shape or size of the specimen, it decreasing markedly with increasing pressure. Thus for a typical medium-hard rubber, μ fell from 0·5 to 0·02 as the load increased from 0·1 to 100 kg cm^{-2}. As the points in fig. 3 show, the following relation is closely obeyed for a wide range of materials and conditions:

$$1/\mu = A(1+Bp/E_0). \qquad \ldots\ldots (1)$$

The value of the coefficient B was almost the same for all materials and averaged 15; the coefficient A ranged from 1 to 8.

The fact that at very high loads the friction force tends to a constant value suggests that the area of real contact ceases to increase. This supposition is supported by results of a simple experiment in which large flat blocks of gelatine were compressed with wavy transparent sheets, representing in exaggerated form the grooves remaining after machining. The outline of the area of real contact was clearly visible, so that the increase of area with load could be determined with reasonable accuracy. The results for three such surfaces are given in fig. 4, where the real contact area/load (a/W) is plotted against p/E_0 so that the curves are comparable with the friction curve of fig. 3. At pressures in excess of $p/E_0 = 0{\cdot}4$ the surfaces were in each case to be in complete contact. There is marked variation in area of contact for a given load between one surface form and another, which suggests that the departure at low loads of the friction curves from the $W^{2/3}$ law could be explained by the fact that real track surfaces are not represented by the multiple spherical asperities on which the law is based.

Effect of Track Material and Lubricant

Load effects were independent of the material of the track and the lubricant, changes merely affecting the value of A. For instance, glass and Perspex gave respectively about 30% and 60% higher friction than steel or rubber; the friction with water was about double that with olive oil, whilst that of dry surfaces was about double again.

With rubber tracks the value of the constant B was found to be one-half that for the same rubber specimen on rigid tracks, a result that is in accordance with a deformation process in which the true area of contact depends on $1/E_1 + 1/E_2$.

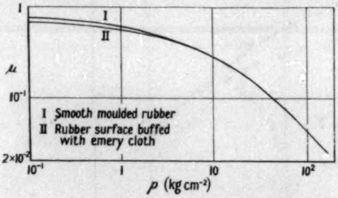


Fig. 5. Variation of coefficient of friction with load, showing effect of roughness of rubber surface. Smooth Perspex track; light mineral oil lubricant, Hycar 75 specimen.

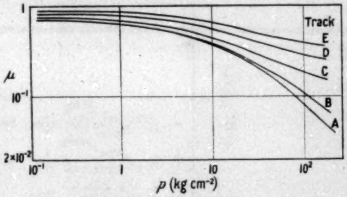


Fig. 6. Variation of coefficient of friction with load, showing effect of roughness of track surface. Smooth specimens of Hycar 90; light mineral oil lubricant; track surfaces as listed in table 2.

Effect of Surface Roughness

With rough rubber specimens sliding on smooth tracks the friction tended to be somewhat less at low loads than with smooth rubber specimens. However, the difference did not exceed 10%, and disappeared at very high loads, as shown in fig. 5.

With rough tracks, on the other hand, the friction increased considerably at all loads. From fig. 6 it can be seen that the excess coefficient of friction of a rough track above that of a smooth track is largely independent of load. With these particular tracks (see table 2), which were prepared by unidirectional abrasion of Perspex blocks with emery cloth of various grades, the friction was higher when the direction of sliding was across the abrasion marks than when parallel to them, and the results given in fig. 7 suggest that the undulation in the direction of sliding has greater influence than the overall height of the asperities.

It is interesting that the track-roughness effect conflicts with the rubber-roughness effect, a phenomenon that has been reported by Roth (1942) for dry rubber at similar speeds.* Only if the effect of roughening were to produce a surface having similar form to surface I in fig. 4 could the real area of contact be greater on rough surfaces, and not only does this seem unlikely, for the tendency disclosed by profilometer traces was in the opposite direction, but such an effect would not persist at very high loads once the specimen was in complete contact with the track.

Table 2. Details of Perspex Tracks

Track	Finish	C.L.A. roughness (microns)	
		(1)	(2)
A	Smooth moulded	0·01	0·01
B	Surface roughened with 00 emery cloth	0·13	0·25
C	,, 0 ,, ,,	0·2	0·38
D	,, 1 ,, ,,	0·38	1·37
E	,, 2 ,, ,,	0·38	1·62

(1) along finishing marks; (2) across finishing marks.

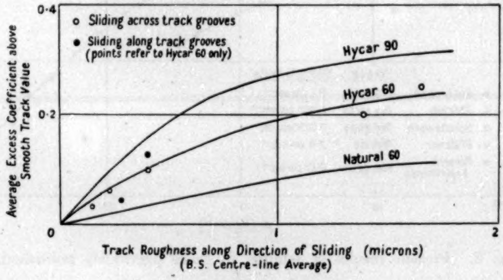


Fig. 7. Relation between excess friction and track roughness. Perspex tracks as listed in table 2 ; light mineral oil lubricant.

It thus seems necessary to assume that, at low sliding speeds at least, track roughness introduces an additional friction force unrelated to contact area, but more or less proportional to load and of sufficient magnitude to mask the slight decrease in contact area that should accompany roughening. Such a force might originate in the energy required to tear out material in the path of track asperities, and this would certainly account for the sensitivity to direction of sliding on grooved surfaces and for the abrasion associated with rough surfaces and high loads. On the other hand, if the rubber were to deform elastically to make way for the track asperities energy would still be lost due to hysteresis accompanying stress reversals in the rubber. It is clear that the present data are insufficient to establish the origin of this force.

---

* Higher sliding speeds did not reverse the present track-roughness effects, as occurs with dry rubber.

Real Frictional Stress

On the assumption that at very high loads the whole area of the specimen is in contact with the track, the average real frictional stress between the surfaces is given by μp or by $E_0/15A$.

The values obtained in this manner vary from 0·33 to 8 kg cm^{-2} for rubber sliding on steel tracks and lubricated with olive oil, the higher stresses corresponding to the harder compounds. Even allowing that the friction might be four times higher with clean dry sliding, the frictional stresses corresponding to these conditions still do not exceed one tenth of the tensile strengths given in table 1. Calculations from estimated properties of the compounds used by Roth, Thirion and Schallamach in their dry friction experiments tend to confirm this.

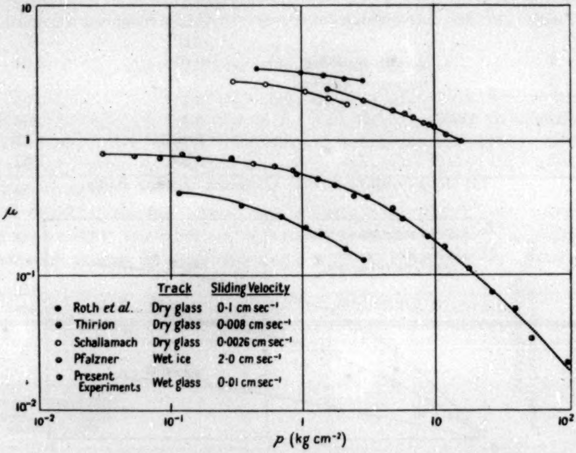


Fig. 8. Present results in relation to those previously published.

§ 3. CONCLUSIONS

1. The load dependence of rubber friction can be satisfactorily explained on the assumption that the surfaces make contact over an area determined by the elastic deformation of surface asperities.

2. At low loads the area of real contact is almost directly proportional to load, whilst at high loads it is independent of load. The resulting friction accords with the empirical relation $1/\mu A = (1 + 15\ p/E_0)$, where A is a constant depending on the nature of the rubber compound, the lubricant and the track.

3. Observed track-roughness effects suggest the existence of a second force unrelated to contact area but dependent on load and on the contour of the track surface. The origin of this force has not been determined, but it appears to be similar to the 'ploughing' term already demonstrated with more rigid materials.

4. The frictional stress based on real contact area does not appear to exceed one tenth of the tensile or tear strength of the rubber even under most severe conditions of dry sliding.

APPENDIX

Figure 8 shows the present curve of friction plotted against load in relation to the results of Roth, of Thirion, of Pfalzner (1950), and of Schallamach.

ACKNOWLEDGMENT

This work was carried out from 1948 to 1950 in the laboratories of the British Hydromechanics Research Association, and the author is indebted to the Council of the Association for permission to publish this account.

REFERENCES

BOWDEN, F. P., *International Conference on Abrasion of Wear*, 1951, *Engineering*, **172,** 724.
LINCOLN, B., 1952, *Brit. J. Appl. Phys.*, **3,** 260.
PFALZNER, P. M., 1950, *Canad. J. Res.* F, **28,** 468.
ROTH, F. L., DRISCOLL, R. L., and HOLT, W. L., 1942, *Bur. Stand. J. Res., Wash.*, **28,** 439.
SCHALLAMACH, A., 1952, *Proc. Phys. Soc.* B, **65,** 657.
THIRION, P., 1946, *Rev. Gen. Caoutch.*, **23,** 101.

The Effect of Temperature on the Mechanical Properties and the Friction of Plastics

By R. F. KING and D. TABOR
Research Laboratory for the Physics and Chemistry of Surfaces,
Department of Physical Chemistry, University of Cambridge

MS. received 23rd April 1953

Abstract. This paper describes a study of the shear strength S, the indentation hardness or yield pressure P and the coefficient of friction μ of polythene, P.T.F.E., Kel-F and Perspex over a temperature range from -100°c up to $+80$°c. In contrast to metals, the plastics show large changes in their strength properties over the temperature range used, and the confusing effect of surface films on the friction has been eliminated. According to the adhesion theory of friction μ is approximately equal to S/P. In these experiments it is found that μ varies closely with the ratio S/P, but the quantitative agreement is good only with polythene. With Perspex and Kel-F the observed values of μ are consistently larger than S/P, while with P.T.F.E. the values of μ are considerably lower. It is concluded that the adhesion theory can explain the friction behaviour of the plastics examined, the apparent discrepancies being due to the enhanced strength of the junctions under the high localized pressures existing at the interface (Perspex and Kel-F) or to structural factors which give a low intrinsic adhesion at the sliding interface (P.T.F.E.).

§ 1. Introduction

The adhesional theory of friction, originally applied by Bowden and his colleagues to metals (Bowden and Tabor 1950) is based primarily on the view that the sliding process involves the shearing of junctions formed by strong adhesion at the actual regions of contact. This may be expressed quantitatively in the following way. When a hard material slides on a softer one under a load W the real area of contact A may be written

$$W = Ap \qquad \ldots\ldots (1)$$

where p is the effective yield pressure of the softer material during sliding. Over this area of intimate contact, junctions are formed of specific shear strength s comparable with that of the softer material itself. The force to shear these junctions, which is essentially the observed frictional resistance, is

$$F = As \qquad \ldots\ldots (2)$$

It is thus possible, as a first approximation, to express the friction in terms of the yield pressure and specific shear strength of the sliding materials.

Recent work has shown that the frictional behaviour of a group of linear polymers can be satisfactorily explained in terms of this theory. Thus autoradiographic experiments (Rabinowicz and Shooter 1952) provide direct evidence for the reality of strong adhesion at the sliding interface, while friction measurements (Shooter and Tabor 1952) yield values of s (eqn. (2)) in good agreement with the bulk shear strength S of the relevant plastic. In general, however, s is found to be greater than S by a factor of 1·2 to 2.

The present paper describes an extension of this work and is concerned mainly with the effect of temperature on the strength properties and friction of the same group of plastics. If we identify the yield pressure p during sliding with the static yield pressure or indentation hardness P, and the shear strength of the junctions s with the bulk shear strength S, we may combine eqns. (1) and (2) to give

$$\mu = \frac{F}{W} = \frac{s}{p} = \frac{S}{P} \qquad \ldots\ldots (3)$$

In the present investigation μ, S and P were measured over a temperature range from $-100°$c to $+80°$c and μ compared with the ratio S/P. A similar investigation for metals has been described by Simon, McMahon and Bowen (1951), but with metals there is a relatively small change in strength properties with temperature and the frictional behaviour is complicated by the influence of surface oxide films. With plastics large changes in S and P are observed over the temperature range used, and by a suitable experimental procedure the effect of surface films on the friction can be eliminated.

§2. EXPERIMENTAL

Four plastics were chosen for this investigation from among those used by Shooter and Tabor. They were

(*a*) Polythene (polyethylene). Most of the experiments were carried out on " Polythene 200 ", the average molecular weight being 15,000. Softening point ~100°c.

(*b*) P.T.F.E. (polytetrafluoroethylene). Produced by sintering the powder above its softening temperature and rapidly quenching the resulting mass. Softening point 327°c.

(*c*) Kel-F (polytrifluorochloroethylene). As supplied by I.C.I. Softening point ~180°c.

(*d*) Perspex (polymethylmethacrylate). Commercial quality. Softening point ~120°c.

Two of these (polythene and Perspex) soften at about 100°c, so the available temperature range was doubled by extending the range of measurement down to $-100°$c. It was in fact found that the more interesting effects were produced below room temperature.

The use of this temperature range made it necessary to carry out the measurements of friction *in vacuo* to avoid the condensation of atmospheric vapours (notably water) on to the specimens. The strength properties are, however, not appreciably affected by surface conditions, and were therefore measured in the atmosphere.

2.1. *Friction*

All experiments were carried out with like pairs of specimens, the upper member of each pair being given a roughly hemispherical shape. A load of about 25 g was used, and the surfaces were made as clean as possible by abrasion under water before mounting them in the apparatus (fig. 1). This consisted of a glass envelope which could be evacuated by a mercury diffusion pump to a pressure, read on an ionization gauge, of less than 10^{-5} mm Hg. The lower friction specimen S_2 was held as rigidly as possible with respect to this envelope by clamps of glass rod, and the upper specimen or slider S_1 was attached to a glass beam J′. By means of a plate P mounted in gimbals and connected through metal bellows to the envelope, the beam could be moved in such a way as to apply normal and tangential forces between the specimens. The magnitude of the forces was measured by means of strain gauges attached to two flat springs G_1 and G_2 between the beam and the plate, and connected in a simple d.c. Wheatstone bridge circuit. The detector in this circuit was a galvanometer of natural frequency high enough to follow the rapid changes in the deflection of the beam, and the overall sensitivity and stability of the arrangement was such that coefficients of friction down to about 0·1 could be measured with reading errors no greater than about 10%.

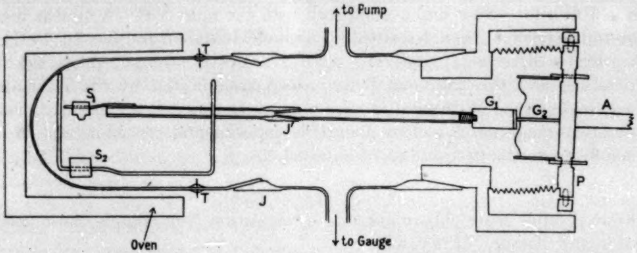


Fig. 1. Diagram of vacuum friction apparatus. S_1, upper surface or slider attached to glass beam. S_2, lower surface held fixed in glass clamp. G_1, G_2, flat spring-steel strips (planes at right angles to each other); strain gauges are attached for determining tangential and normal forces. P, brass plate supporting upper slider, attached to envelope through metal bellows. J, J', ground joints for assembling apparatus. T, tungsten wire seals for setting up a cleaning discharge or for connecting to thermocouple. The envelope is mounted vertically for immersion in liquid air.

Cleaning of specimens.

After the specimens had been mounted in this apparatus, the coefficient of friction μ between them was measured. The envelope was then heated gently with the specimens apart until a good vacuum and a reproducible coefficient of friction were obtained after cooling. The values of μ obtained before and after this cleaning by heating *in vacuo* are shown in table 1.

Table 1. Effect of Cleaning on Friction of Plastics.
(Friction measurements at 20°c)

	Coefficient of Friction		
Plastic	Abraded under water	Heated *in vacuo*	Nature of sliding
Polythene	0·5	0·80	Smooth
Fluon	0·1	0·10	Smooth
Kel-F	0·4	0·55	Stick slip
Perspex	0·5–0·6	0·62	Stick slip

It will be seen that the friction is increased little, if at all, by this treatment. This is surprising, in view of the very great effect of surface films on the friction of both metals and non-metals reported by Bowden and Young (1950), and suggests that plastics may perhaps be free from such surface films. It is, however, possible that relatively gentle heating is insufficient to remove contamination from the specimens. This seemed particularly likely in the case of P.T.F.E. which was shown by Shooter and Tabor to have a coefficient of friction only one third of that expected on the adhesion theory. Specimens of P.T.F.E. were therefore subjected to the more drastic cleaning treatment of heating to 300°c *in vacuo* for several hours. The coefficient of friction measured at room temperature remained equal to 0·1, suggesting that even this anomalously low value is probably not due to contamination.

A few similar drastic heating experiments with the other plastics also gave little change in the coefficient of friction. It would seem that with these plastics surface films do not appreciably affect the friction or are easily removed.

The effect of temperature.

Most of the measurements of friction were made in the temperature range +20°c to −100°c on specimens which had been cleaned in the way described above. The envelope of the vacuum apparatus was immersed in a Dewar flask containing liquid air, and values of μ were obtained at frequent intervals as the specimens cooled. Their temperature was measured at the same time by means of a thermocouple embedded just below the surface of the lower specimen. This was calibrated at −183°c, 0°c and +100°c. When equilibrium had been reached, after about an hour, the Dewar flask was removed, and the observations of friction and temperature continued.

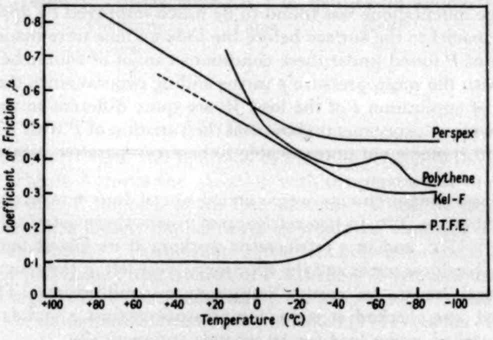


Fig. 2. Effect of temperature on the friction of clean plastics.

During the cooling part of the cycle the friction varied reproducibly with the temperature. During the warming part of the cycle, however, a temporary fall in friction to about $\mu = 0{\cdot}2$ was observed with all the plastics except P.T.F.E., and at the same time the pressure in the apparatus rose to about 10^{-3} mm Hg. This effect was probably due to ice evaporating from the envelope and condensing on to the specimens. For this reason, in considering the variation of μ with temperature, only the results obtained during the cooling part of the cycle have been used. These are given for all the plastics in fig. 2.

The rise in the friction of polythene near −60°c was observed with several specimens of two different mean molecular weights, and seems to be genuine. At about the same temperature, the nature of the sliding process changes, marked 'stick-slip' motion developing in contrast to the smooth sliding observed at room temperatures. The values of μ found for this plastic and for P.T.F.E. at room temperature are rather higher than those given by Shooter and Tabor, but the discrepancy can be partly accounted for in terms of the increase in friction with decreasing load found by these workers.

The sliding of specimens of Kel-F was of the 'stick-slip' type over the whole temperature range. The friction continued to increase above room temperature, though no accurate measurements were made in this range. Perspex also showed a steady variation in μ, with no change to smooth sliding, over the whole temperature range from $+100°$c to $-60°$c.

2.2. *Static Yield Pressure or Hardness*

The conditions chosen as standard for the measurements of this property were as follows. A Vickers 136° diamond pyramid was loaded with 1 kg and mounted on a beam in such a way that it could be lowered on to the specimen. The square indentation produced was examined microscopically and the mean length d of its diagonal measured. The static yield pressure or hardness P which is the load divided by the *projected* area of the indentation (Tabor 1951) was then calculated from the expression $P=2000W/d^2$ where P is in kg mm^{-2}, W is the load in grammes (in this case 1000) and d is in microns. The use of a pyramidal indenter avoids difficulties due to the elastic recovery of the indentation, since the diagonals remain practically unaltered in length during recovery. The visibility of the indentations was found to be much improved by applying a thin coat of black enamel to the surface before the indentations were made.

The value of P found under these conditions cannot of course be assumed to be identical with the mean pressure p in the sliding process since the magnitude and the time of application t of the load W are quite different in the two cases. However, subsidiary experiments show that the variation of P with W and t is not very rapid, and it seems not unreasonable to base comparative measurements on P rather than on p.

The hardness measurements were carried out at four temperatures: in the laboratory at about $+20°$c, in two refrigerated rooms thermostatically controlled to $-5°$c and $-17°$c, and in a refrigerator working at its lowest temperature of $-40°$c. The hardness tester and the specimens were left in the room or refrigerator for several hours, to ensure temperature equilibrium. The portable apparatus used was checked at room temperature against a standard hardness tester which applied a 1 kg load for 10 seconds automatically.

Results.

The hardness values were all determined at a load of 1 kg and for loading times of 10, 20 and 40 seconds. In general it was found that, over this time range, the effect of loading time on the hardness was masked by the errors in measurement. For the smaller hardness values (i.e. larger indentations) the overall error is not greater than about 10%; for the larger hardness values the error is somewhat greater. Average values for the four plastics are given in table 2.

2.3. *Shear Strength*

In these measurements a simple apparatus was used, similar to that of Shooter and Tabor, in which a cylindrical specimen of the plastic was sheared between two metal blocks. This does not give a state of homogeneous shear stress but provides a fair approximation to the shear conditions existing within the interfacial junctions.

All the shear strength measurements were carried out in the refrigerator. The apparatus was loaded with a specimen of plastic, left to reach temperature equilibrium, and operated from outside to break the specimen. The force necessary to do this was read on a spring balance, and the value of S was calculated by dividing the force by twice the cross-sectional area of the specimen, since shearing takes place across two planes. The results for the four plastics are summarized in table 3, the experimental error being of the order of 7 to 10%.

The strength of Perspex was also measured at +80°c, and was found to be 5·4 kg mm^{-2}.

Table 2. Static Yield Pressure or Hardness P of Plastics (in kg mm^{-2})

Temp. (°c)	Polythene	Fluon	Kel-F	Perspex
+20	1·3	5·3	11·5	23
− 5	2·6	5·9	16·4	†
−17	3·9	7·0	28	†
−40	6·1	7·5	*	38

* Impression too small for measurement.

† No measurement made. The hardness of Perspex was, however, measured at +80°c, and found to be 18 kg mm^{-2}.

Table 3. Shear Strength S of Plastics (in kg mm^{-2})

Temp. (°c)	Polythene	Fluon	Kel-F	Perspex
+20	0·8	1·8	4·5	5·2
−17	1·7	2·5	5·0	
−40	2·1	3·6	6·6	5·6

§ 3. DISCUSSION

These data on strength and hardness can now be related to the measurements of friction. Figure 3 shows the variation with temperature of the coefficient of friction μ and of S/P for each of the four plastics. It will be seen that the two quantities vary in a closely similar way, as would be expected if the adhesion theory were applicable. The absolute agreement between μ and S/P is, however, good only for polythene : the friction of P.T.F.E. is lower, and that of the other two plastics higher, that that predicted by the simple adhesion theory.

Simon, McMahon and Bowen have published (1951) results similar to these for some metals over a temperature range of 4·2°K to 600°K. They also find, in general, a parallel variation of μ and S/P, the former always being greater, usually by a factor of two or three. As one interpretation of these results one may suggest a relation between μ and S/P of the form

$$\mu=\mu_0+kS/P \qquad \ldots\ldots (4)$$

where μ_0 is a contribution to the friction due to some factor other than adhesion (e.g. electrostatic forces, surface roughness) and k is a constant of order of magnitude unity but generally higher.

The data of fig. 3 have therefore been plotted in fig. 4 as a graph of μ against S/P. The broken line represents the relation $\mu=S/P$ expected from the simple adhesion theory. It is seen that the values for polythene fall close to the theoretical line, Perspex and Kel-F are appreciably above it, and P.T.F.E. below it. If we attempt to fit the results into a relation of the type given in eqn. (4) it is found that the intercept μ_0 on the axis of μ is positive for Kel-F, zero for Perspex and negative for polythene and P.T.F.E. In view of the wide limits of error, indicated on fig. 4 by the shaded areas, and the difficulty of giving any physical significance to negative values of μ_0 it seems better to assume that μ_0 is always zero and to represent the results by the best straight lines through the origin. This has been done in fig. 5, and the corresponding values of k in the assumed relation $\mu = kS/P$ are: Perspex 2·4, Kel-F 1·7, polythene 1·1, P.T.F.E. 0·3.

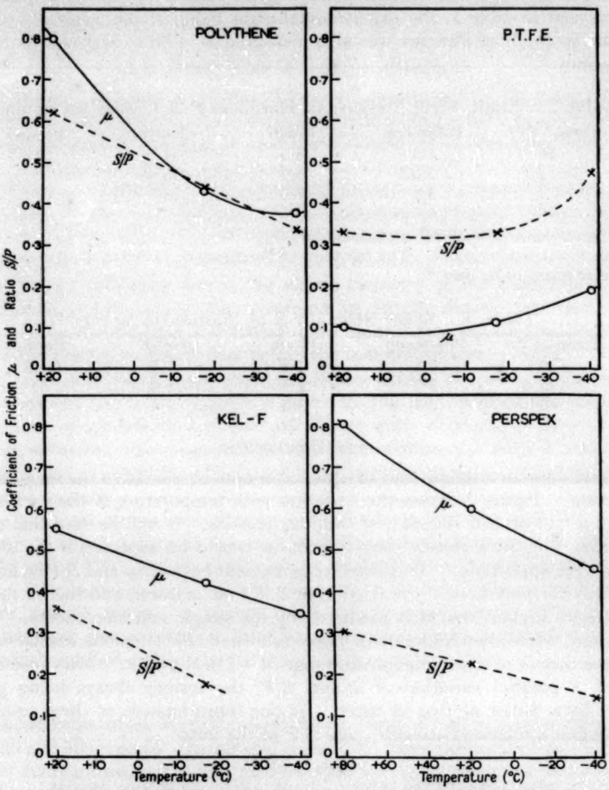


Fig. 3. Comparison of μ and S/P as a function of temperature.

The experiments by Shooter and Tabor at heavier loads gave ratios k of about the same order as those quoted here, and it is interesting to consider the reason for this. In the friction measurements described here it was not possible, at the low loads used, to measure the width of the track produced during sliding. However, in the experiments by Shooter and Tabor the track widths could be observed, and from these the mean pressure between the surfaces while they were actually sliding could be calculated. These agreed with the static hardness values to within about 10%. Thus the discrepancy between μ and S/P is due primarily to the difference between the specific shear strength S of the material in bulk and the specific shear strength s of the material at the sliding interface. This effect is observed to an even more striking extent with markedly brittle materials such as rock salt, glass and sulphur. In experiments with these materials (King 1952, King and Tabor, to be published) it is found that the interfacial

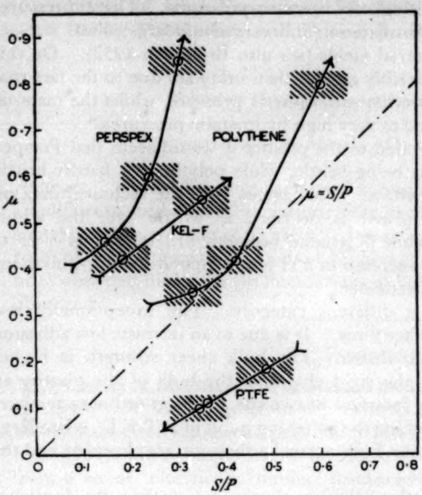


Fig. 4. Relation between μ and S/P for clean plastics. Arrows indicate direction of increasing temperature. Shaded regions indicate experimental error.

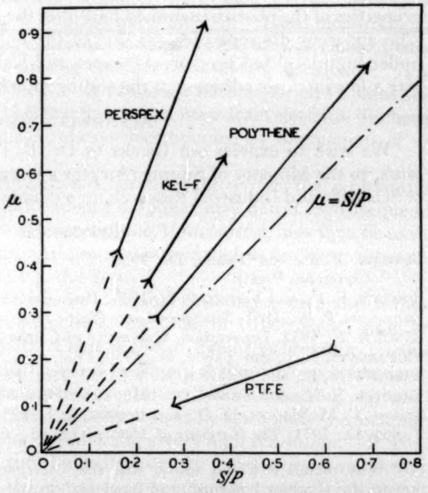


Fig. 5. Relation between μ and S/P. Results of fig. 4 redrawn on assumption that $\mu = kS/P$.

junctions are at least ten times stronger in shear than the material in bulk. Detailed examination shows that this is primarily due to the fact that the friction junctions are subjected to a complex state of stress equivalent to a shear stress on which are superposed very high hydrostatic pressures. These pressures prevent premature fracture of the material, and very much larger shear stresses can be withstood before the material yields (see also Bridgman 1952). On this view, therefore, values of *k* appreciably greater than unity are due to the fact that the bulk shear strength is measured at atmospheric pressure, whilst the material at the sliding interface is subjected to very high hydrostatic pressures.*

If this explanation can be applied to the plastics it would seem that Perspex and Kel-F may be considered as being brittle, while polythene is hardly brittle enough to show any anomalous effect. This agrees with the technical data on Perspex and polythene. The former is decribed as brittle over the whole temperature range considered, while polythene becomes brittle only at -25°c. The slight tendency towards an increase in *k* at low temperatures for polythene may be connected with this embrittlement.

P.T.F.E. clearly falls into a different category. The exceptionally low friction ($k \simeq 0{\cdot}3$) is not due to surface films. It is due to an intrinsic low adhesion between the P.T.F.E. molecular chains. The bulk shear strength is higher because of interlocking between the rigid chains in the bulk of the plastic, an effect which is not possible at the interface of two specimens at ordinary temperatures. At temperatures approaching the softening point of P.T.F.E., when large scale flow at the interface can occur, high normal adhesions are observed and the friction can rise to very high values.

It is evident, therefore, that the adhesion theory can explain the frictional properties of the plastics examined here over the whole temperature range applied. The apparent discrepancies are due to the enhanced strength of the junctions under high localized pressures (Perspex and Kel-F) or to structural factors which give a low intrinsic adhesion at the sliding interface (P.T.F.E.)

ACKNOWLEDGMENTS

We wish to express our thanks to Dr. F. P. Bowden for his interest in the work, to the Ministry of Supply (Air) for a research grant and to the Department of Scientific and Industrial Research for a maintenance grant to one of us (R.F.K.).

REFERENCES

BOWDEN, F. P., and TABOR, D., 1950, *The Friction and Lubrication of Solids* (Oxford : Clarendon Press).
BOWDEN, F. P., and YOUNG, J. E., 1950, *Proc. Roy. Soc.* A, **208,** 444.
BRIDGMAN, P. W., 1952, *Studies in Large Plastic Flow and Fracture* (London: McGraw-Hill).
KING, R. F., 1952, *Dissertation,* University of Cambridge.
MCFARLANE, J. S., and TABOR, D., 1950, *Proc. Roy. Soc.* A, **202,** 244.
RABINOWICZ, E., and SHOOTER, K. V., 1952, *Proc. Phys. Soc.* B, **65,** 671.
SHOOTER, K. V., and TABOR, D., 1952, *Proc. Phys. Soc.* B, **65,** 661.
SIMON, I., MCMAHON, H. O., and BOWEN, R. J., 1951, *J. Appl. Phys.*, **22,** 177.
TABOR, D., 1951, *The Hardness of Metals* (Oxford : Clarendon Press).

---

* With metals it would appear that the discrepancy between μ and S/P is primarily due to the combined normal and tangential stresses at the regions of intimate contact; these lead to further plastic flow and growth of the intermetallic junctions (McFarlane and Tabor 1950). Such junction growth appears to be negligible with plastics except at temperatures near their softening point.

Study of High-Frequency Fluctuations from Light Sources using Phototubes and Tuned Radio-Frequency Amplifiers

By K. LANDECKER and B. J. ROBINSON
School of Physics, University of Sydney, Australia

Communicated by H. Messel; MS. received 15th April 1953

Abstract. Phototubes have been used in conjunction with tuned radio-frequency amplifiers to study light fluctuations from discharge tubes and from the Sun (sun spots) up to frequencies of 95 Mc/s. In particular the spatial distribution of some types of oscillations along the axis of discharges in helium and mercury vapour have been obtained but so far no fluctuations of light from the 'quiet' Sun have been observed.

§ 1. Introduction

In a series of investigations of ionized media carried out in this School (Bailey 1948, 1950, 1951) numerous experiments were performed on oscillations and noise in electrical discharges (Bailey and Landecker 1950, Middlehurst 1950, Fyfe 1951). Early experiments included also the comparison of the light flux from a set of 'electrically similar' discharge tubes in which a phototube, light modulator (chopper) and associated amplifier were used (V. A. Bailey, K. Landecker, R. F. Mitchell and J. A. Roberts, unpublished). These measurements were sometimes spoiled through spurious readings which, as was realized later, were due to the discharge tubes passing through some state of oscillation. The technique was therefore abandoned. It has since been reported (Cobine and Callagher 1947, in particular p. 114) that fluctuation noise from electrical discharge tubes is, at least at low frequencies, always accompanied by fluctuations of the light emitted by these tubes.

In the course of these experiments it occurred to the writers that a phototube used in conjunction with a radio-frequency amplifier may detect high-frequency fluctuations from light sources. This technique, if successful, may then be useful in the investigation of oscillations in laboratory sources such as discharge tubes and even in extra terrestrial bodies as e.g. the Sun.

It is well known that noise arises in electrical discharges, with and without magnetic fields at least up to frequencies of the order of 10 000 Mc/s (Mumford 1949, B. J. Robinson, unpublished). There is also an extensive literature on noise from luminous (and non-luminous) celestial bodies up to about the same frequency.

In gas discharges it is important to localize the oscillations occurring, if possible without the disturbing intervention of probes. In radio astronomy the technique to be described appears promising in providing a very high resolution. It would also be of interest to correlate in time fluctuations at the frequencies of visible light and at radio frequencies when the signal components have passed through the ionosphere or through interstellar matter. This would allow one to determine the dispersion and supply information about the density in space of the intervening matter.

We have so far only been successful in designing tuned phototube amplifiers working at frequencies up to 95 Mc/s using commercial phototubes. It is the purpose of this note to report on our initial experiments with these devices and to give a brief description of the apparatus used.

§ 2. THE SENSITIVITY TO BE EXPECTED FROM A TUNED PHOTOTUBE AMPLIFIER AT HIGH FREQUENCIES

Consider the performance of a phototube connected across the input (tuned radio-frequency) circuit of a radio receiver in the manner indicated in fig. 1. It is clear that one should select for this purpose a high vacuum type of phototube. Then it is easy to estimate the sensitivity of this arrangement : the resonant impedance $L\omega Q$ of the tuned circuit may be made at least 10^5 ohms over the entire range of frequencies in which a coil and condenser combination is used. The internal (dynamic) resistance of the phototube itself in the range of anode voltages of interest is of the order of 10^8 ohms and need not be considered. Further it should be possible to detect a variation of (noise or 'signal') amplitude of about one microvolt developed across the circuit. This corresponds to a photocurrent signal amplitude of 10^{-11} A. At full illumination of the photo surface (e.g. 10 μA average photocurrent with the RCA 919 tube) we should therefore be able to detect a fluctuation of about 1/10000 of one per cent of the average photocurrent and therefore of the average illumination. This estimate of the order of magnitude of the sensitivity to be expected was confirmed by experiments described below.

Since phototubes are used here in a rather ususual way it seems desirable to make the following remarks : we are not directly concerned with the limiting sensitivity of the tube but assume that a certain minimum of illumination is available from the source. Then a corresponding shot noise current will flow, whatever the nature of the source, and the amplifier may or may not raise this to a detectable level. The quantity we wish to measure is either an excess of noise above shot noise or a periodic oscillation superimposed on the shot noise, both being special properties associated with certain sources.

We restricted ourselves in this work to commercially produced phototubes. Transit-time effects were avoided by using these tubes with the highest permissible anode voltage. We have investigated the use of photomultipliers for sources of low intensity but here the difficulties seem to be much greater.

It is hoped to consider in a later publication the theory of the processes entering into the present technique.

§ 3. PHOTOTUBE EQUIPMENT IN THE FREQUENCY RANGE 20 c/s TO 95 Mc/s

We had at our disposal a communication receiver covering the frequency range 500 kc/s to 25 Mc/s and a prewar v.h.f. superheterodyne receiver (Model T.F. 554 made by Marconi–Ekco) covering the range 20 to 300 Mc/s.* Both receivers were fitted with output meters.

(*a*) *Frequency range* 500 *kc/s to* 25 *Mc/s*.

The mode of connecting the phototube (RCA type 919) to the input circuit of the communication receiver is shown in the circuit diagram (fig. 1). The addition of the phototube does not affect the alignment or the tuning of the receiver in any way. It was convenient to mount the tube inside the metal receiver cabinet. The battery and photocurrent meter were placed in a separate screened box. For the adjustment of the illumination a polaroid shutter and a set of filters are used. At the full illumination and with the gain controls fully advanced the shot noise from the phototube is strongly audible on all bands and can be read on the output meter.

---

* Kindly lent to us by Dr. J. L. Pawsey of the Commonwealth Scientific and Industrial Research Organization, Sydney. This receiver is at present used on another project.

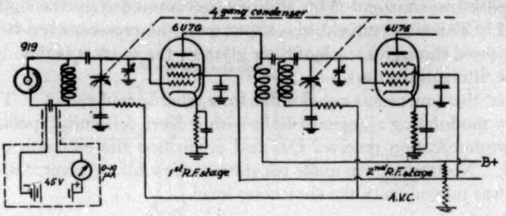

4 gang condenser
6U7G
6U7G
919
45V
0-10 μA
1st R.F. stage
2nd R.F. stage
B+
A.V.C.

Fig. 1. Mode of connection of phototube RCA type 919 to the communication receiver.

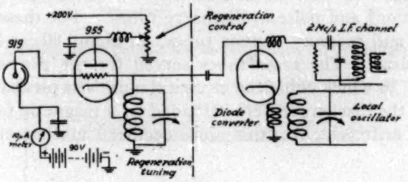

+200V
Regeneration control
955
919
2 Mc/s I.F. channel
Diode converter
Local oscillator
10 μA meter
90V
Regeneration tuning

Fig. 2 (*a*). Input circuit of the v.h.f. receiver with the phototube connected.

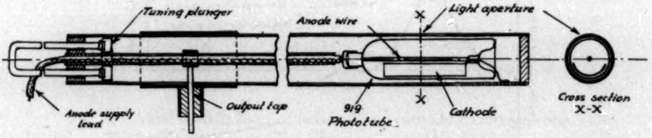

Tuning plunger
Anode wire
Light aperture
X
X
Anode supply lead
Output tap
919 Phototube
Cathode
Cross section X-X

Fig. 2 (*b*). An alternative input circuit for the phototube using a coaxial line.

(*b*) *Frequency range* 20–95 Mc/s.

The v.h.f. receiver used a diode frequency changer and no radio-frequency amplification and was not nearly as sensitive as the communication receiver. The shot noise of the phototube could not be detected on any frequency band. Since shot noise was at that stage our main criterion of the sensitivity, and of the usefulness of the equipment for the applications we had in mind, reaction was applied to the input circuit in the manner indicated in fig. 2 (*a*). The reaction valve is an acorn triode (type 955). In this way we succeeded in extending the useful range of this equipment to about 95 Mc/s which is just near the frequency at which solar noise is recorded by the Radio Astronomy Group of the Commonwealth Scientific and Industrial Research Organization, in Sydney (97 Mc/s). The tuning of the receiver was surprisingly easy because the regenerative stage could be used as an oscillator. An alternative input circuit using a coaxial line is shown in fig. 2 (*b*).

These receivers together with low-frequency equipment available in this department covered the range 20 c/s to 95 Mc/s.

§ 4. MEASUREMENT OF THE SENSITIVITY OF THE EQUIPMENT

Tests were carried out on a number of light sources mainly for the purpose of finding a convenient source for calibration. A strong response was obtained by illuminating the phototube in the communication receiver with the light from the gap of a spark transmitter. This disappeared immediately the light beam was interrupted. There was an audible output from the receiver even when the light beam had passed through a stack of filter glasses: the spark was then barely visible through the filters in the dark.

However, this source was not suitable for a reliable calibration. This has been achieved by modulating a beam of light with a Kerr cell and a special oscillator. For the communication receiver this test confirmed the estimate of sensitivity made in § 2. No attempt was made to calibrate the v.h.f. receiver: at all times the sensitivity was judged from the shot noise level.

§ 5. EXPERIMENTS ON THE LIGHT FROM ELECTRICAL DISCHARGES IN GASES

These investigations were undertaken mainly for the purpose of localizing electrical oscillations and noise in discharge tubes. For these experiments hot cathode helium and mercury vapour tubes, all about 80 cm long and of 5 cm diameter, were used. The same tubes served for the previous investigation, referred to in § 1, in which enhanced electrical noise was picked up from external electrodes when the discharges were pervaded by a magnetic field parallel to the direction of the drift velocity; this noise occurred in characteristic frequency bands.

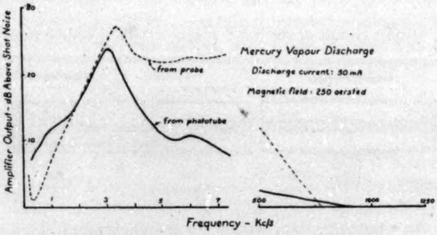


Fig. 3. Noise spectrum picked up by a probe and by the phototube from mercury vapour discharge with axial magnetic field.

The same noise could be detected when the light from the discharge was focused on to the phototubes of our apparatus. The noise spectra were determined in the ranges 200 to 7500 c/s and 0·5 to 2 Mc/s and agreed strikingly with those picked up directly by means of external electrodes and the same amplifiers (receivers) (fig. 3). The higher frequency bands observed previously were above the range of our receivers.

A screen with a slit 4 mm wide was placed in front of the discharge so that the light from a narrow transverse section could be focused on to the photo surface (fig. 4). The optical axis of this arrangement could be moved bodily along the whole length of the discharge. Figure 5 is a schematic diagram of the noise amplitude plotted against tube length obtained at all frequencies in the range of the equipment. The shape of the plot is the same for both the gases investigated.

This makes it appear likely that the disturbance is mainly localized near discontinuities in the neighbourhood of the cathode and to a much lesser degree at the anode. This is in agreement with observations of other workers (Merrill and Webb 1939, Emeléus and Armstrong 1949, Armstrong *et al.* 1951, Martin and Woods 1952), utilizing probe methods. The main body of the positive column when observable does not appear to provide directly a notable part of this type of noise. Reversal of the direction of the magnetic field does not noticeably alter the appearance of the plot. It is important to point out that the spatial variation of noise amplitude is not due to a variation in the brightness of the discharge because the photo current (illumination) was kept constant.

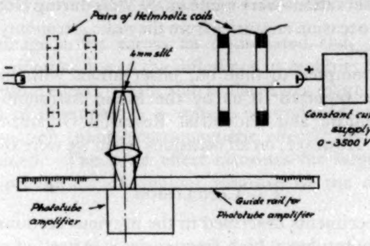

Pairs of Helmholtz coils
4mm slit
−
+
Constant current supply 0 - 3500 V
Phototube amplifier
Guide rail for Phototube amplifier

Fig. 4. Optical arrangement for investigations on discharge tubes.

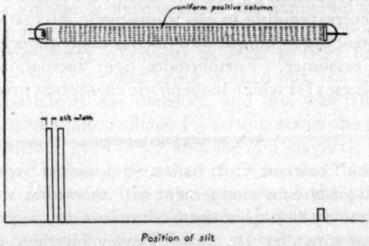


Fig. 5. Axial distribution of noise in hot cathode discharge tube.

§ 6. EXPERIMENTS ON THE IMAGE OF THE SUN'S DISC

For these experiments the 6 in. telescope of this department was used to project an image of the Sun of about 30 cm diameter on to a plane containing the photosurface. The light was directed to the phototube by means of a mirror in such a way that the entire image could be scanned, the effective aperture being about 1 cm^2. The illumination was again controlled by a set of filters but these were not used purposely to select any particular part of the spectrum. Particular attention was paid to sunspots.

We realize that this experiment is in principle entirely different from the experiments described in §§4 and 5 in as much as the dimensions of the source are extremely large compared with the wavelengths investigated. However, since there is to our knowledge no conclusive reason why light fluctuations should not occur, correlated with the electrical fluctuations received at various frequencies, we have carried out some preliminary experiments. Observations were made during May 1951 with the communication receiver. On no occasion was there a noticeable excess of noise above the shot noise level produced by a steady source over the frequency range 500 kc/s to 25 Mc/s. After completion of the high frequency equipment similar observations were made on 90 Mc/s during November 1951 and May 1952. On each occasion enquiries from the radio astronomy workers showed that the Sun was 'quiet'.

We have since attempted to time our observations with the onset of noise outbursts (storms) as reported to us by the Radio Astronomy Group of the Commonwealth Scientific and Industrial Research Organization in Sydney working at 97 Mc/s. However, on all occasions when we were free to observe the sky was overcast.

§ 7. CONCLUSION

Regarding the experiments described in the previous sections we believe that the performance of phototubes at high frequencies is in itself of some interest and we hope that the simple technique described will be of use in clarifying some aspects of electrical discharges.

As far as the astronomical aspects mentioned are concerned it is clear that, if light fluctuations corresponding to electromagnetic fluctuations can be detected at all, the resolution of the technique described is vastly superior to even the largest radio telescope in existence. Furthermore, light fluctuations may obviously be picked up at frequencies for which ionospheric absorption precludes the reception of radio waves.

ACKNOWLEDGMENTS

We wish to thank Professor V. A. Bailey, Professor of Experimental Physics in this School, for advice and encouragement with this work. We are also obliged to the Radio Astronomy Group of the Commonwealth Scientific and Industrial Research Organization under Dr. J. L. Pawsey for their co-operation and to Mr. M. H. Brennan for assistance with the design and calibration of apparatus. The work has benefited from a grant for equipment from the Electrical Research Board of Australia.

REFERENCES

ARMSTRONG, E. B., EMELÉUS, K. G., and NEILL, T. R., 1951, *Proc. Roy. Irish Acad.* A, **54**, 291.

BAILEY, V. A., 1948, *Aust. J. Sci. Res.* A, **1**, 351; 1950, *Phys. Rev.*, **78**, 428; 1951, *Ibid.*, **83**, 439.

BAILEY, V. A., and LANDECKER, K., 1950, *Nature, Lond.*, **166**, 259.

COBINE, J. D., and CALLAGHER, C. J., 1947, *J. Appl. Phys.*, **18**, 110.

EMELÉUS, K. G., and ARMSTRONG, E. B., 1949, *Proc. Instn. Elect. Engrs.*, **96**, Pt. III, 390.

FYFE, D. F., 1951, *Thesis*, University of Sydney.

MARTIN, H., and WOODS, H. A., 1952, *Proc. Phys. Soc.* B, **65**, 281.

MERRILL, H. J., and WEBB, H. W., 1939, *Phys. Rev.*, **55**, 1191.

MIDDLEHURST, J., 1950, *Thesis*, University of Sydney.

MUMFORD, W. W., 1949, *Bell Syst. Tech. J.*, **28**, 608.

Photoelectromagnetic and Photodiffusion Effects in Germanium

By T. S. MOSS, L. PINCHERLE and A. M. WOODWARD
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 9th March 1953

Abstract. A voltage arises across an illuminated slab of germanium in the direction of illumination due to the difference in mobility between electrons and holes (photodiffusion effect). If the slab is placed in a magnetic field perpendicular to the direction of illumination, a voltage is produced at right angles to both field and illumination (photoelectromagnetic effect); voltages as great as 0·5 v have been obtained. The latter effect saturates for large illumination and its limiting value provides a convenient measure of the surface recombination velocity.

§1. INTRODUCTION

We have observed that if a slab of semiconductor, illuminated on one of its faces, is placed in a magnetic field parallel to the illuminated surface, a voltage is developed in the semiconductor in a direction perpendicular to the magnetic field and to the normal to the illuminated surface (photoelectromagnetic effect). Voltages as great as 0·5 v have been obtained. While the theory of this effect was being worked out, the existence of another effect (photodiffusion effect) was deduced, and this was later detected experimentally. It consists of a photovoltage V_{D} arising across the plate perpendicular to the illuminated surface, even in the absence of a magnetic field, and is caused by the difference in mobility between electrons and holes: V_{D} is in fact proportional to this difference. On the other hand the photoelectromagnetic voltage V_{H} which is proportional to the magnetic induction and may be considered as the Hall effect produced by the diffusion current, or a particular case of the Suhl effect (Shockley 1950) depends, for small signals, on the sum of the mobilities. If recombination of holes and electrons occurs mainly in the bulk of the material, V_{D} is large and V_{H} is small; if on the other hand recombination occurs mainly on the surface V_{D} is negligible and V_{H} is large.*

§2. MATHEMATICAL FORMULATION

The theory has been developed for the arrangement illustrated in fig. 1.

The symbols used throughout are: I, illumination in quanta m^{-2} sec^{-1} absorbed on xz face; κ, absorption coefficient of semiconductor (m^{-1}); n_0, number of electrons m^{-3} in unilluminated slab; n_0+n, equilibrium density of electrons in illuminated slab (m^{-3}); p, equilibrium density of holes in illuminated slab (m^{-3}); e, absolute value of electronic charge in coulomb; D, diffusion constant of holes (m^2 sec^{-1}); μ, hole mobility (m^2 volt^{-1} sec^{-1}); b, electron mobility/hole mobility; ϵ, dielectric constant (F m^{-1}); r, bulk recombination factor (m^{3} sec^{-1}); eJ_{h}, hole current (A m^{-2}); eJ_{e}, electron current (A m^{-2}); E, electric field (V m^{-1}); $B \equiv B_z$, magnetic induction (weber m^{-2}); s, surface recombination velocity (m sec^{-1}).

---

* After this paper had been written, a paper was published (Aigrain and Bulliard 1953) describing the photoelectromagnetic effect, which had apparently been observed before (Kikoin and Noskov 1934, Frenkel 1935) in CuO at liquid air temperatures. Also a voltage between the dark and the illuminated side of a semiconducting plate had been observed by Dember (1931).

The existence of surface states and thus of a space charge layer at the free surface of the semiconductor will be neglected throughout this paper. Thus the boundary conditions for the electric field are that its normal component must vanish on any surface. This approximation neglects completely any change of surface potential that may be caused by illumination (Brattain 1951). As these effects may be important, the formulae derived for the photodiffusion effect may not be easily verified experimentally.

The semiconductor, assumed to be n-type, of thickness t in the y direction is taken infinite in the z direction so that the problem is a two-dimensional one in the plane of the paper. It is assumed that, in addition to the normal bulk recombination of holes and electrons, there is a surface recombination obeying the law $J = sp$, where J is the flow into the surface. In the absence of a magnetic field, the hole and electron distributions are independent of the x coordinate, and for a small applied magnetic induction B the variation with x is neglected. Then the total current in the y direction must be zero since it vanishes on the surfaces, and we may put $J_{\mathrm{h}y} = -J_{\mathrm{e}y} = J$.

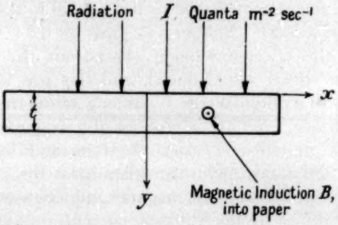


Fig. 1.

Also if the quantum efficiency is unity, then for I quanta m^{-2} sec^{-1} absorbed on the xz face, $\kappa I \mathrm{e}^{-\kappa y} dy$ holes and electrons m^{-2} sec^{-1} are separated in a thickness dy of the material at a depth y.

Then the equations giving p, n, E_y and J are ($B = 0$),

$$J = -D\frac{dp}{dy} + \mu p E_y \qquad \ldots\ldots(1)$$

$$-J = bD\frac{dn}{dy} + b\mu(n_0 + n)E_y \qquad \ldots\ldots(2)$$

$$\frac{dJ}{dy} = I\mathrm{e}^{-\kappa y} - r(n_0 + n)p \qquad \ldots\ldots(3)$$

$$n - p = -\frac{\epsilon}{e}\frac{dE_y}{dy}. \qquad \ldots\ldots(4)$$

The treatment of the general case would be involved, therefore only the limiting cases of small signal and large signal will be considered.

2.1. *Small Signal Theory, no Magnetic Field*

Suppose that I is small, so that $p, n \ll n_0$; then, as the conduction term in (2) is of the same order as the diffusion term, the conduction term in (1) may be neglected. Putting $\tau = 1/rn_0$ = lifetime of holes in the bulk of the material (about 200 μsec for intrinsic specimens at room temperature) equations (1) and (3) give

$$p = p_0\, e^{-\kappa y} + p_1 \sinh \frac{y}{l} + p_2 \cosh \frac{y}{l} \qquad \ldots\ldots(5)$$

$$J = \kappa D p_0\, e^{-\kappa y} - p_1 \frac{D}{l} \cosh \frac{y}{l} - p_2 \frac{D}{l} \sinh \frac{y}{l} \qquad \ldots\ldots(6)$$

where $p_0 = \kappa I \tau / 1 - \kappa^2 l^2$, $l^2 = D\tau$.

In germanium, $D = 4 \times 10^{-3}$ m^2 sec^{-1} at ordinary temperatures and, with $\tau \sim 100\,\mu$sec, l is of the order of a fraction of a millimetre. In all cases of practical importance κ is at least 10^6 m^{-1}, thus $\kappa^2 l^2 \gg 1$ and

$$p_0 = I\tau/\kappa l^2 = -I/\kappa D. \qquad \ldots\ldots(7)$$

The boundary conditions are: $y = 0$, $J = -sp$; $y = t$, $J = sp$.

If t is larger than a few microns, terms containing the factor $e^{-\kappa t}$ may be neglected; and, since s rarely exceeds 100 m sec^{-1}, it may be neglected with respect to κD; then introducing the dimensionless quantity $\alpha = D/ls$ which, in most practical cases, has a value between 0·1 and 1,

$$p = \frac{I}{s} \frac{\sinh \{(t-y)/l\} + \alpha \cosh \{(t-y)/l\}}{(1+\alpha^2) \sinh (t/l) + 2\alpha \cosh (t/l)} - \frac{I}{\kappa D}\, e^{-\kappa y} \qquad \ldots\ldots(8)$$

$$J = I\alpha \frac{\cosh \{(t-y)/l\} + \alpha \sinh \{(t-y)/l\}}{(1+\alpha^2) \sinh (t/l) + 2\alpha \cosh (t/l)} - I e^{-\kappa y}. \qquad \ldots\ldots(9)$$

In both expressions the second term is negligible, except within a few microns of $y = 0$.

Eliminating n from (2) and (4) we find the differential equation for E_y, the electric field which ensures that the total current vanishes:

$$\frac{d^2 E_y}{dy^2} - \frac{e\mu n_0}{\epsilon D} E_y = \frac{e}{\epsilon b D}\left(bD \frac{dp}{dy} + J\right), \qquad \ldots\ldots(10)$$

p and J being given by eqns. (8) and (9).

For germanium (of ordinary purity) $(\epsilon D / e\mu n_0)^{1/2}$ is 10^{-7} m or less, thus the 'transient' part of the solution of (10), which gives the behaviour of E_y near $y = 0$ and $y = t$ (at which points $E_y = 0$), may be ignored for the purpose of calculating the potential difference between the two sides of the plate. We are left with the 'steady state' solution, which is in fact the solution for zero charge ($p = n$). This condition is approximately true in the interior of the slab to within about $(\epsilon D / e\mu n_0)^{1/2}$ of the surface. Then

$$E_y = \frac{I\alpha(b-1)}{b\mu n_0 A}\left[\cosh \frac{t-y}{l} + \alpha \sinh \frac{t-y}{l}\right] \quad \text{where} \quad A = (1+\alpha^2) \sinh \frac{t}{l} + 2\alpha \cosh \frac{t}{l}$$

and thus

$$V_D = -\int_0^t E_y\, dy = -\frac{I\alpha(b-1)l}{b\mu n_0 A}\left[\sinh \frac{t}{l} + \alpha\left(\cosh \frac{t}{l} - 1\right)\right]. \qquad \ldots\ldots(11)$$

2.2. *Small Signal Theory in the Presence of a Magnetic Field*

The application of a small magnetic field in the z direction causes a negligible change in J (if $\mu^2B^2 \ll 1$). Then, neglecting the small variation of n and p with x,

$$J_{hx} = \mu p E_x + \mu BJ,\ J_{ex} = b\mu(n_0+n)E_x + b\mu BJ.$$

At the boundaries, the total current must be zero. Since the variation with x is neglected, put $J_{hx}+J_{ex}=0$ for all x. Thus:

$$E_x = -\frac{1+b}{b}\,\frac{BJ}{n_0+n+p/b} \simeq -\frac{1+b}{b}\,\frac{BJ}{n_0},\ \text{since } p,\ n \ll n_0$$

and if the distance between the probes is d, $V_H \simeq \{(1+b)/b\}BJd/n_0$. Since J varies across the plate, V_H is a function of y. However, the approximations introduced are probably not valid near the front surface, where all the absorption takes place. At the back surface $(y=t)$, $J(t)=I\alpha/A$,

and
$$V_H(t) = \frac{1+b}{b}\,\frac{Bd}{n_0}\,\frac{I\alpha}{(1+\alpha)^2\sinh(t/l)+2\alpha\cosh(t/l)}. \qquad \ldots\ldots(12)$$

For $\alpha \to 0$ (total surface recombination)

$$V_D \to 0;\ V_H(t) \to \frac{1+b}{b}\,\frac{BId}{n_0\kappa l\sinh(t/l)}$$

(the latter expression does not follow from (12), as the approximation $s \ll \kappa D$ is not valid, but may be easily derived from the basic equations),

for $\alpha \to \infty$ (no surface recombination)

$$V_D \to -\frac{b-1}{b}\,\frac{Il}{\mu n_0}\,\frac{\cosh(t/l)-1}{\sinh(t/l)};\ V_H \to 0.$$

A rough estimate of the short circuit current through the sample may be derived very simply as follows: if μB is the Hall angle for holes, the holes will move a distance μBl_h in the direction of the electrodes, and the electrons a distance $b\mu Bl_e$, where l_h and l_e are the effective diffusion lengths for holes and electrons when surface recombination is taken into account.

Hence
$$i_s = e\mu BcI(l_h + bl_e) \qquad \ldots\ldots(13)$$

where c is the width of the sample.

2.3. *Large Signal Theory*

This approximation assumes that the specimen is so pure, or that I is so large, that $p,\ n \gg n_0$. To obtain an approximate solution of eqns. (1) to (4) under these conditions bulk recombination is neglected, and it is assumed that the absorption coefficient is so large that the illumination penetrates no further than the surface. Then part of the holes and electrons produced by illumination will be destroyed by recombination at the surface $y=0$, the remainder flowing forward into the slab.

With these assumptions, eqns. (1) to (4) reduce to

$$\left.\begin{array}{l} J = -D\dfrac{dp}{dy} + \mu p E_y, \quad -J = bD\dfrac{dn}{dy} + b\mu n E_y \\ \dfrac{dE_y}{dy} = \dfrac{e}{\epsilon}(p-n), \quad J = \text{constant} < I. \\ \text{The boundary conditions are} \\ E_y = 0 \text{ at } y=0 \text{ and } y=t, \quad I-J = sp \text{ at } y=0, \quad J = sp \text{ at } y=t. \end{array}\right\} \ldots\ldots(14)$$

Then
$$J\left(1-\frac{1}{b}\right)=-D\frac{d}{dy}(p-n)+\mu E_y(p+n) \qquad \ldots\ldots(15)$$
$$J\left(1+\frac{1}{b}\right)=-D\frac{d}{dy}(p+n)+\mu E_y(p-n). \qquad \ldots\ldots(16)$$

Integrating (16) gives
$$p+n=p(0)+n(0)-\frac{J}{D}\left(1+\frac{1}{b}\right)y+\frac{\mu\epsilon}{2e}E_y^{\,2}. \qquad \ldots\ldots(17)$$

In particular, for $y=t$,
$$\frac{J}{s}+n(t)=p(0)+n(0)-\frac{Jt}{D}\left(1+\frac{1}{b}\right). \qquad \ldots\ldots(18)$$

Since $d(p-n)/dy$ is small except near $y=0$ and t, we may neglect it in the interior of the slab. Also the last term in (17) is small.

Then in the interior
$$E_y=\frac{J(1-b^{-1})}{\mu(p+n)}=-\frac{b-1}{b+1}\frac{D}{\mu}\frac{d(p+n)/dy}{p+n}. \qquad \ldots\ldots(19)$$

Hence
$$V_{\mathrm{D}}=-\int_0^t E_y\,dy=\frac{b-1}{b+1}\frac{D}{\mu}\ln\frac{(J/s)+n(t)}{p(0)+n(0)}. \qquad \ldots\ldots(20)$$

The quantities $n(t)$, $p(0)$ and $n(0)$ are determined from the more exact solution for E_y, satisfying the boundary conditions at $y=0$ and t.

The differential equation satisfied by E_y, neglecting only the non-linear term in (17), is
$$\frac{d^2E_y}{dy^2}=-\frac{eJ(1+b^{-1})}{\epsilon D}+\frac{E_y}{L^2}$$

where
$$\frac{1}{L^2}=\frac{\mu e}{\epsilon D}\left[p(0)+n(0)-\frac{J}{D}(1+b^{-1})y\right]. \qquad \ldots\ldots(21)$$

Assuming that the variation of L across the slab is small and that $L\ll t$, we get
$$E_y\simeq\frac{eJ(1-b^{-1})}{\epsilon D}\left[L^2-L(0)^2\exp(-y/L)-L(t)^2\exp\{(y-t)/L\}\right] \qquad \ldots\ldots(22)$$

Then at $y=0$,
$$p(0)-n(0)\simeq\frac{J}{D}(1-b^{-1})L(0)=\frac{J}{D}(1-b^{-1})\left(\frac{\epsilon D}{\mu e}\right)^{1/2}[p(0)+n(0)]^{-1/2} \qquad \ldots\ldots(23)$$

and at $y=t$,
$$\frac{J}{s}-n(t)\simeq-\frac{J}{D}(1-b^{-1})L(t)=-\frac{J}{D}(1-b^{-1})\left(\frac{\epsilon D}{\mu e}\right)^{1/2}\left[\frac{J}{s}+n(t)\right]^{-1/2}. \qquad \ldots\ldots(24)$$

For s small, $n(t)\to J/s$. In fact putting $\Delta=n(t)-J/s$, eqn. (24) gives
$$\Delta^2=\epsilon sJ(1-b^{-1})^2/2De\mu\simeq 1{\cdot}9\times10^{11}sJ.$$

Hence the approximation is good as long as $s^3\ll10^{-11}J$. This will be true under all practical conditions for large signals. Hence, using (18), (21), (23) and (24)
$$n(t)\simeq J/s \qquad \ldots\ldots(25)$$
$$p(0)\simeq n(0)\simeq\frac{Jt}{2D}\left(1+\frac{1}{b}\right)+\frac{J}{s}. \qquad \ldots\ldots(26)$$

The densities and the field are now determined at all points as functions of J, which may be related to I through the boundary condition $I-J=sp(0)$; thus
$$J=\frac{I}{2+st(1+b^{-1})/2D}. \qquad \ldots\ldots(27)$$

Finally,

$$V_{\mathrm{D}}=-\frac{b-1}{b+1}\frac{D}{\mu}\ln\left[1+\frac{st}{2D}\left(1+\frac{1}{b}\right)\right] \qquad \ldots\ldots(28)$$

and

$$E_x(0+)=-\frac{Bs}{1+st(1+b^{-1})/2D},\quad E_x(t)=-Bs. \qquad \ldots\ldots(29)$$

The assumption that the variation of L across the slab is small is not a good one for all practical values of s, but improves as s decreases. However numerical integrations for the case $s\rightarrow\infty$, where the assumption is worst, give a correction factor of only 0·61 in the value of $n(t)$. We are therefore encouraged to think that the final formulae are sufficiently good for all practical values of s, tending to give values of s in excess by not more than 40%.

It is interesting that both the diffusion and magnetic effects are in this case independent of the illumination; both provide in principle a method for the measurement of the surface recombination velocity, but only the latter is simple and practicable. The second of the equations (29) may be deduced very simply as follows: when only the electrons produced by illumination are of importance, they will move into the surface at $y=t$ with a velocity equal to the surface recombination velocity s. They are thus subjected on the surface to a lateral force eBs which, as no current flows in the x direction, must be compensated by an equal and opposite force due to an electric field component E_x. Hence $E_x=-Bs$.

§3. EXPERIMENTAL METHODS AND RESULTS

3.1. *Specimens and Measuring Conditions*

The specimens used were in the form of rectangular rods 1–2 cm long and about 1 mm × 0·5 mm cross section. The resistivity of the germanium used was either 0·014 Ωm or 0·40 Ωm. For the lower resistivity material the carrier concentration would be approximately $1\cdot3\times10^{21}$ electrons m^{-3}, whilst the higher resistance specimens would be near the intrinsic range with approximately 10^{19} holes m^{-3} and 4×10^{19} electrons m^{-3}. Soldered (ohmic) contacts were used throughout. It was found that these did not produce any appreciable photovoltages.

White light illumination from tungsten filament lamps was used except when measuring the spectral distribution of sensitivity. The lamps were operated at their normal working temperatures (~2500°c) and the rate of emission of useful quanta (i.e. those with $h\nu>0\cdot75$ ev) found from the usual radiation formulae. The power radiated was checked by a calibrated thermopile at some particular range from the lamp, and the inverse square law then used to find the intensity at other ranges. The values quoted for the illumination are corrected for reflection losses.

The magnetic field was provided by a d.c. electromagnet with pole faces 3·8 cm diameter. Fields up to 1·2 weber m^{-2} could be obtained with a 1·5 cm gap.

The photoelectromagnetic effect was measured on microammeters or spot galvanometers, the resistance of the measuring circuit being made either large or small compared with the specimen resistance to obtain open circuit voltages and short circuit currents respectively.

3.2. *Photoelectromagnetic Effect. Linearity of Photocurrent–Photovoltage Relation*

Observations were made of the photocurrent for various values of r and R using the circuit shown in fig. 2, the illumination being kept constant. With $R=\infty$, r was varied from 0 to 11 kΩ. A plot of the inverse of the current against resistance is shown by curve A. The linearity of the results shows the system to be ohmic with a source voltage of 13·7 mv and an impedance of 480 Ω.

The results of curve B were obtained by varying R with $r=0$. The plot of inverse current against inverse resistance again shows the system to be ohmic with a source voltage of 13·7 mv.

3.3. *Photoelectromagnetic Effect. Linearity of Short Circuit Current with Intensity and Magnetic Field*

Figure 3 shows the short circuit current plotted against the intensity of irradiation expressed in terms of quanta $\text{m}^{-2}\,\text{sec}^{-1}$. Comparison with the straight line drawn at an angle of 45° with the axes shows that the relation is substantially linear over the whole range, although the specimen resistance decreased 12 : 1 under the strongest illumination.

The short circuit current is linear with magnetic field up to $B \simeq 0{\cdot}5$ weber m^{-2} for higher values of B the current increases less rapidly than B. Assuming the current to be reduced by a factor $1+(\mu B)^2$, where μ is an effective mobility, as in the magneto-resistance effect, a value of $\mu = 0{\cdot}45\ \text{m}^2\,\text{v}^{-1}\,\text{sec}^{-1}$ is obtained. This is higher than the values found by Hall effect measurements, but is in reasonable agreement with values obtained from magneto-resistance measurements (see Moss 1952).

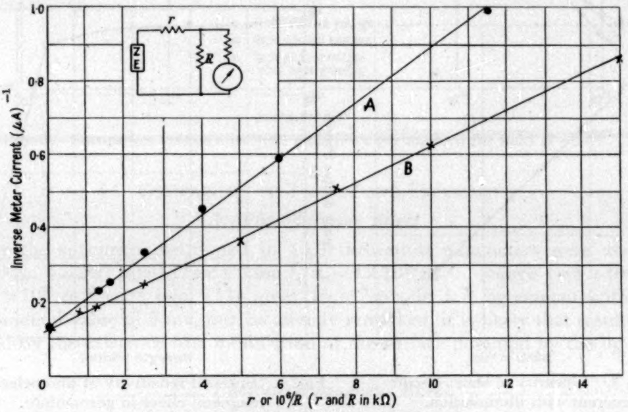


Fig. 2. Linearity of photoelectromagnetic effect.

3.4. *Photoelectromagnetic Effect. Spectral Distribution of Sensitivity*

This was measured over the wavelength range 0·7–2μ using a rock-salt prism double monochromator. The radiation was interrupted at 800 c/s and the signal from the sample amplified at this frequency. As shown by fig. 4, the quantum efficiency is roughly constant between 0·7 and 1·6μ, but falls rapidly at longer wavelengths. From the curve the sensitivity is seen to have fallen to half its maximum value at $\lambda_{1/2} = 1{\cdot}63\,\mu$. This wavelength corresponds to an energy of 0·76 ev, in good agreement with results from conventional photoconductivity measurements and with the accepted forbidden energy gap of germanium (Moss 1952).

3.5. *Measurement of the Photodiffusion Voltage*

After the presence of the photodiffusion effect had been deduced an attempt was made to detect the effect experimentally. As it was thought that the use of metal contacts would produce photovoltages which might well mask the diffusion voltage, they were avoided and a condenser arrangement used with interrupted illumination. The slab of germanium was illuminated through a thin sheet of mica on which was painted a serious of conducting lines (of platinum), the lines being on the side of the mica away from the germanium. The back face of the germanium slab was insulated from a flat metal electrode by a second sheet of mica. The signals were amplified in a high gain amplifier tuned to the interruption frequency.

Signals many times amplifier noise were obtained and the actual magnitude of the photovoltage was found by direct calibration by injecting a known signal in series with the lead to the back electrode. An r.m.s. voltage of 2·2 mv was obtained for a very pure specimen for an intensity of 10^{21} quanta m^{-2} sec^{-1}. With allowance for wave form factors this is equivalent to about 5 mv peak to peak voltage. The experimental arrangement, however, does not allow discrimination between the photodiffusion effect and the effect of illumination on the surface potential. The latter would be stronger on the front surface than on the back surface.

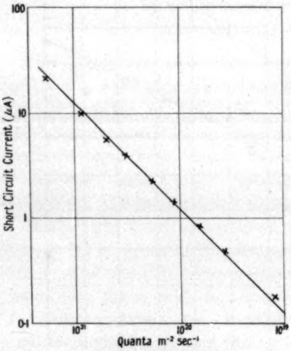


Fig. 3. Linearity of short-circuit current with illumination.

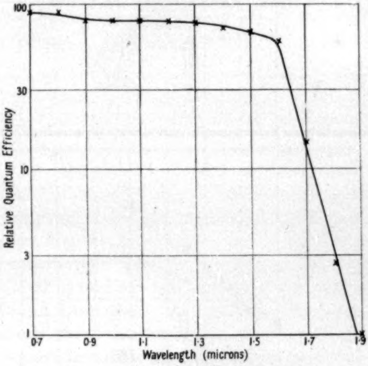


Fig. 4. Spectral sensitivity of photoelectromagnetic effect in germanium.

3.6. *Measurement of Surface Recombination Velocity*

At high intensities of illumination eqn. (29) shows that the electric field should reach a saturation value which is directly proportional to the surface recombination velocity s. The curves of fig. 5 show that such behaviour does occur, saturation taking place with about 10^{22} quanta m^{-2} sec^{-1} or more for the high resistivity germanium. For the initial state of the sample (where it was etched, but not very thoroughly) the saturation voltage developed on the back surface given by fig. 5 is 56 mv, corresponding to 3·1 v m^{-1}. As the magnetic field used was $B=0{\cdot}49$, the equation $E_x(t)=Bs$ gives $s=6{\cdot}4$ m sec^{-1}. Measurement on the front surface gave $E_x(0)=1{\cdot}9$ v m^{-1} so that from the first equation (29) we again find $s \sim 6\ \mathrm{m\,sec}^{-1}$. The effect of grinding the surface with fine carborundum is shown by the upper curve of fig. 5. The saturation photovoltage has risen to 0·23 v, corresponding to $s = 26\ \mathrm{m\,sec}^{-1}$. The maximum value of s observed after various grinding treatments was $33\ \mathrm{m\,sec}^{-1}$.

These values are in general agreement with those quoted by Shockley (1950) as deduced from lifetime measurements on injected carriers.

This experiment thus provides a particularly simple and direct method of measuring the surface recombination velocity.

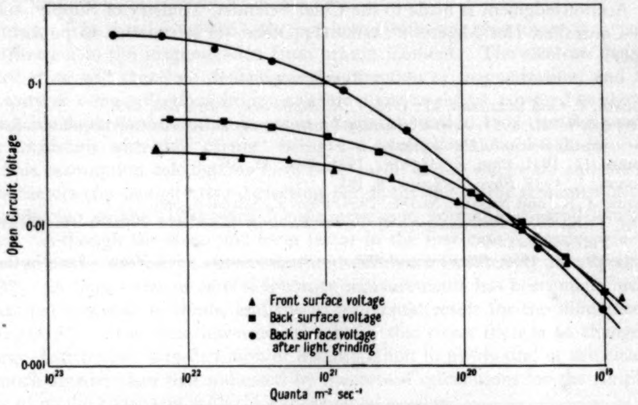


Fig. 5. Saturation voltages in high resistivity germanium for field $B = 0{\cdot}49$.

§4. Comparison of Theory and Experiment

4.1. *Photodiffusion Effect*

For the specimen mentioned in §3.5 theoretical parameters were $\alpha = 0{\cdot}2$, $l = 10^{-3}$ m, $b = 2{\cdot}1$, $\mu = 0{\cdot}17\ \mathrm{m}^2\,\mathrm{v}^{-1}\,\mathrm{sec}^{-1}$, $n_0 = 4 \times 10^{19}\ \mathrm{m}^{-3}$. Since $t = 5 \times 10^{-4}$ m and $I = 10^{21}\ \mathrm{m}^{-2}\,\mathrm{sec}^{-1}$ eqn. (11) gives $V_{\mathrm{D}} = 7$ mv, in fair agreement with the experimental value of 5 mv, but, as already remarked, it is likely that results are vitiated by the existence of a modulation of the surface potential by the light.

4.2. *Photoelectromagnetic Effect*

For the low resistivity sample n_0 was $10^{21}\ \mathrm{m}^{-3}$ and an intensity of $7{\cdot}5 \times 10^{20}\ \mathrm{m}^{-2}\,\mathrm{sec}^{-1}$ was used; $B = 0{\cdot}55$ weber m^{-2}, $d \sim 0{\cdot}02$ m; $t/l \sim 1{\cdot}5$ and $\alpha \sim 1{\cdot}0$. Thus formula (12) gives $V_{\mathrm{H}} = 1{\cdot}4$ mv on the back surface. The observed value on the front surface was 1·44 mv; that on the back surface was not measured but would be somewhat higher.

For the high resistivity sample n_0 was $4 \times 10^{19}\ \mathrm{m}^{-3}$. At an intensity of $5 \times 10^{19}\ \mathrm{m}^{-2}\,\mathrm{sec}^{-1}$ the observed voltage on either surface was 10 mv; $B = 0{\cdot}49$ weber m^{-2}, $d \sim 0{\cdot}02$ m, $t/l \sim 0{\cdot}6$ and $\alpha \sim 1{\cdot}0$; thus formula (12) gives $V_{\mathrm{H}} = 5$ mv. Similar results are obtained for other experimental points of fig. 5. The agreement is not so good, but probably in these conditions the small signal theory is not a good approximation. For the same sample formula (13) gives a short circuit current of $(10^{-26} I)$ A in good agreement with the results given in fig. 3.

As experimental conditions are very different from the ideal case considered theoretically, deviations from the simple formulae given may be expected, due to the breaking down of one or more of the simplifying assumptions. End effects are likely to be particularly important.

ACKNOWLEDGMENTS

Acknowledgment is made to the Chief Scientist, Ministry of Supply, and to the Controller, Her Majesty's Stationery Office, for permission to publish this paper.

REFERENCES

AIGRAIN, P., and BULLIARD, H., 1953, *C.R. Acad. Sci., Paris*, **236,** 595.

BRATTAIN, W. H., 1951, *Semi-Conducting Materials*, ed. H. K. Henisch (London : Butterworths Scientific Publications), p. 37.

DEMBER, H., 1931, *Phys. Z.*, **32,** 554; 1932, *Ibid.*, **33,** 207.

FRENKEL, J., 1935, *Phys. Z. Sowjet*, **8,** 185.

KIKOIN, I. K., and NOSKOV, M. M., 1934, *Phys. Z. Sowjet*, **5,** 586.

MOSS, T. S., 1952, *Photoconductivity in the Elements* (London : Butterworths Scientific Publications).

SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

On the Nature of Ferromagnetism in Pyrrhotite

By K. ALEXOPOULOS and A. THEODOSSIOU
Department of Physics, University of Athens

Communicated by E. C. Stoner; *MS. received 5th May* 1953

Abstract. The *g*-value of 0·63 for pyrrhotite indicates that there is a large contribution to the magnetization from orbital moment. The electron density distribution will therefore depend on the direction of magnetization, and the intensity of x-ray reflections from a suitable plane might be expected to show a detectable dependence on this direction. Inglis showed in 1934 that the *g*-value was consistent with the 'carrier' being a d-electron with $m_l = +2$, $m_s = -\frac{1}{2}$. On this assumption calculations have been made of the electronic and atomic form factors (for one effective d-electron per atom) and of the structure factors for reflection on the (110) plane for magnetization normal and parallel to this plane. Although the electronic form factor in the first case is nearly twice as great as in the second, the final calculated difference in intensity is only about 0·83%. A long series of careful intensity measurements has been made, using ionization chamber methods, and the experimental result for the difference is 0·30 ± 0·15%. It is thus shown experimentally that either there is no change in charge distribution with direction of magnetization in pyrrhotite, or the change is much smaller than that indicated by theoretical calculations for the simplest type of model consistent with the experimental *g*-value.

§ 1. Introduction

The current theory of ferromagnetism assumes that at temperatures well below the Curie point each domain is spontaneously magnetized almost to the saturation value, with almost all magnetic carriers oriented in the same direction. The nature of the magnetic carriers is not unique. In most ferromagnetics, gyromagnetic experiments give for the Landé factor a value $g = 2$, which shows that the crystalline field completely quenches the orbital moments, so that only spin moments are spontaneously oriented, thus contributing to magnetization. In pyrrhotite (Fe_7S_8) however, Coeterier (1933, 1935) found a value $g = 0·63$; this is an indication that in this ferromagnetic the orbital moment is not quenched. The observed value ($\sim\frac{2}{3}$) was attributed by him to Fe-atoms in a $^2P_{1/2}$ state. On the other hand Inglis (1934) showed that Coeterier's results could be explained if it is assumed that the magnetic moment of each magnetic carrier is due to a single d-electron with a strong spin–orbit antiparallel coupling. Whereas without spontaneous magnetization all orientational states of the resulting moment $j = 3/2$ are equally populated, in the case of saturation the degeneracy is removed and only the lowest state is occupied, corresponding to an alignment of all vectors. This can be described as all elementary magnets—in this case d-electrons—being oriented in the direction of magnetization. As d-electrons have a density distribution that is not spherically symmetrical, the possibility would exist of changing the electronic density distribution in the crystal by changing the direction of magnetization. This deduction can, in principle, be experimentally verified with x-rays.

The possibility that change of direction of magnetization might result in a change of x-ray reflection intensities must often have been considered, but the amount of published work on the subject seems to be very small. Over thirty years ago Compton and Rognley (1920) made measurements of the intensity of x-rays reflected from a crystal of magnetite in the demagnetized state and when magnetized to about one third of the saturation value. Within the limits of accuracy of the experiments—it was estimated that a difference of about 1% could have been detected—no change was observed. An interesting discussion is given of the significance of this nul result in the light of the current theoretical outlook (see also Compton 1921). More recently, Stephens (1938) has made measurements on the effect of change of magnetization on the intensity of x-ray reflection from crystals of pyrrhotite, but with no conclusive results.

Let us consider a pencil of rays falling on a crystal at a suitable Bragg angle. The intensity of reflected rays will depend, *inter alia*, on the atomic form factors and thus upon the density distribution.

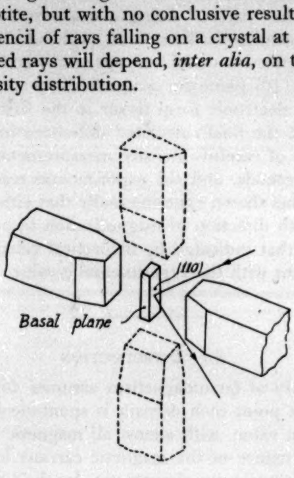


Fig. 1.

If the density distribution is spherically symmetrical (as is the case with atoms with s-symmetry or atoms with d-symmetry but equally distributed over all *m*-values) the atomic form factor will depend only on the scattering angle and will have the values computed by Hartree (1933). In pyrrhotite there is no such symmetry and therefore the form factor will not only depend on the scattering angle but will also vary from domain to domain depending on its direction of magnetization.

If the pyrrhotite is introduced into a saturating magnetic field the form factor will have the same value in all domains and will depend upon the direction of the field. The form factor will be largest for a field in a direction bisecting the angle between the primary and the reflected beam, i.e. for a field normal to the reflecting lattice plane (fig. 1, full lines). This can easily be seen in the simplified case of an electron moving on a circular orbit. Under the field just mentioned, the orbit will become parallel to the reflecting lattice plane, and the form factor will have a maximum value equal to unity. Now if the field is turned through 90°, the plane of the orbit will become normal to the reflecting lattice plane and the form factor will acquire a minimum value.

Actually, the circular orbits of the classical picture have to be replaced by an electronic distribution of 3d symmetry, but as will be shown in §2 the above results are still qualitatively valid.

Changes in the value of the form factor affect the intensity of the reflected x-ray beam and in §3 we describe an experimental arrangement suitable for investigating this change. In §4 we survey the experimental results and discuss the conclusions that can be drawn from them.

Before proceeding further it should be noted that Inglis's assumption cannot apply to all the iron atoms. The magnetic moment of an atom with $l=2$ and $m_s=-\frac{1}{2}$ is $1\,\mu_B$. If we assume that all Fe atoms in pyrrhotite have this moment, the computed saturation magnetization is three times larger than that observed (Weiss and Forrer 1929). It is thus only every third Fe atom to which the above can apply, and the rest must be assumed to have no magnetic moment whatever.

There is another possible way of explaining the observed small value of magnetization: pyrrhotite is considered by Néel (1952) to be a ferrimagnetic, consisting of two unequal sublattices. The first is normally populated with Fe atoms whereas in the second many sites for Fe atoms are vacant. The Fe atoms are probably in an ionic state (see Hirone and Tsuya 1951) corresponding to the formula $Fe_5{}^{2+}Fe_2{}^{3+}S_8$. The total magnetization is equal to the difference of the magnetization of the two sublattices. The Fe^{2+} and Fe^{3+} ions must have such magnetic moments and must be distributed in such a way as to give the correct value of magnetization.

§ 2. COMPUTATION OF THE FORM FACTOR

The intensity of a beam of x-rays, reflected on lattice plane (hkl) of a crystal, depends on the square of the structure factor S which is given by the formula

$$S=\sum_{\kappa} F_\kappa \exp \{i2\pi(hx+ky+lz)\},$$

where F is the atomic form factor of atom κ, and x, y, z give its position in the unit cell.

It is clear that any change of the values of F—due for example to the application of a magnetic field—will result in a change of S and therefore in a change in the intensity of the reflected beam. F depends on the distribution of the electronic density round the nucleus. If we consider the density as a sum of terms, each due to a separate electron we may write the atomic form factor as a sum of electronic form factors $F=\Sigma_z f$, f being given by

$$f=\int \sigma \exp \{(i4\pi h/\lambda) \sin \alpha\}\, dv \qquad \ldots\ldots(1)$$

where σ is the electronic density and h is the distance of the elementary volume dv from the reflecting plane. If, for a given electron, σ is spherically symmetrical, as in the case of an s-electron, the result of the integration will be exclusively a function of glancing angle α. If, on the contrary, σ is not symmetrical, as in the case of a d-electron, it will be given by a product of three functions R, Θ, Φ, and the value of f will therefore depend on the orientation of the axis of d-symmetry relative to the incoming and reflected rays. This question of orientation does not affect cases where a shell is full. But also in cases where a shell is not full, distribution over all states with different *m* quantum numbers will result in spherical symmetry. In this way Hartree has calculated the values of f for different types of electrons, and of F for most elements.

In the case of pyrrhotite in a strong magnetic field the problem acquires a different aspect, as in a given domain all magnetic carriers are assumed to have the same orientation. If we identify each magnetic carrier with a d-electron, all Fe atoms will exist in the state of highest *m* quantum number (equal to 2). The value of f and F will depend now upon the orientation of the field relative to the incoming and reflected rays. We shall therefore have to calculate the values of F from case to case. As the exact density distribution is not known we will approximate by making certain simplifying assumptions. The Fe atom is considered to consist of a spherically symmetrical core and a single hydrogen-like 3d electron, that is supposed to be the exclusive carrier of magnetic moment. We thus write $F = F_{\mathrm{core}} + f_{3\mathrm{d}}$. For the core we take the values F_{H} given by Hartree for the Fe atom and subtract the form factor $f_{3\mathrm{d},\,\mathrm{H}}$ of one *spherically symmetrical* 3d electron (i.e. distributed equally upon all *m* numbers) as given by the same author: $F_{\mathrm{core}} = F_{\mathrm{H}} - f_{3\mathrm{d},\,\mathrm{H}}$. The value of $f_{3\mathrm{d}}$ was calculated from formula (1) by using hydrogen-like eigenfunctions (Pauling and Wilson 1935) with $n=3$, $l=2$, $m=2$ for the electronic density σ. For the effective atomic number Z^* we subtract from $Z=26$ the dimensional shielding factor (Pauling and Sherman 1932) $s=16{\cdot}4$ and obtain $Z^*=9{\cdot}6$. The above eigenfunctions have rotational symmetry around a so-called z-axis. The evaluation of integral (1) depends upon the orientation of this axis relative to the incoming and outgoing rays. We will consider three cases: (*a*) the applied magnetic field is normal to the reflecting plane, (*b*) it is parallel, and (*c*) there is no applied magnetic field. In cases (*a*) and (*b*) the external field must be parallel to the basal plane because pyrrhotite displays ferromagnetism only in this plane.

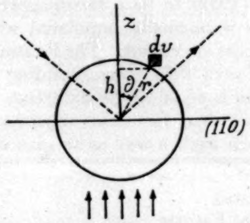


Fig. 2.

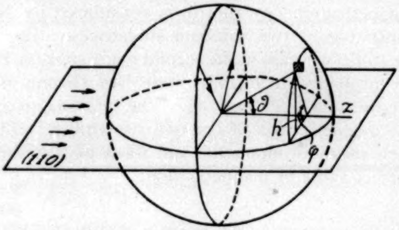


Fig. 3.

A preliminary calculation (Alexopoulos 1942) carried out by the same general method, but using a model with a circular orbit instead of hydrogen-like density functions, showed that the lattice plane (110) is the most suitable plane for experiments, i.e. for a given change of F this plane gives the largest change in reflected radiation. All subsequent calculations have been carried out for values of $\sin\alpha/\lambda$ corresponding to this lattice plane. The directions of the field in relation to the reflecting crystal are shown in fig. 1.

(*a*) *Field normal to lattice plane* (110). In this case the z-axis is normal to the reflecting plane (fig. 2) and hence h can be written $h = r \cos\theta$. Graphical integration (Alexopoulos and Theodossiou 1950) of eqn. (1) using the aforementioned hydrogen-like eigenfunctions gives $f_{3d} = 0{\cdot}696$.

(*b*) *Field parallel to lattice plane* (110). In fig. 3 the z-axis is drawn parallel to lattice plane (110) and normal to the plane defined by the incoming and the reflected beams, corresponding to the arrangement chosen in the experiments described below. In this case we get $h = r \sin\theta \sin\phi$ and the graphical integration gives $f_{3d} = 0{\cdot}371$.

(*c*) *Without external field*. Pyrrhotite is an anisotropic crystal of the hexagonal system and shows ferromagnetism only in the basal plane. In this plane three directions (α, β, γ) of easiest magnetization have been observed, at angular intervals of 60°. We assume that in each domain the magnetization and therefore the z-axis coincides with one of these directions. It further remains to relate these to the three crystallographic axes (1, 2, 3). This has not been done by the investigators of the magnetic qualities of pyrrhotite but for reasons of symmetry it is plausible to assume that in hexagonal crystals we will find one of the two following cases: the z-axis either coincides with (case 1), or is exactly normal to (case 2) the crystallographic axes 1, 2 and 3 (fig. 4). The pyrrhotite crystal consists of three kinds of domains α, β, γ equally distributed relatively to the three crystallographic axes 1, 2, 3. The electronic form factor f_{3d} will thus be the mean of three values f_α, f_β and f_γ, each of which belongs to one kind of domain.

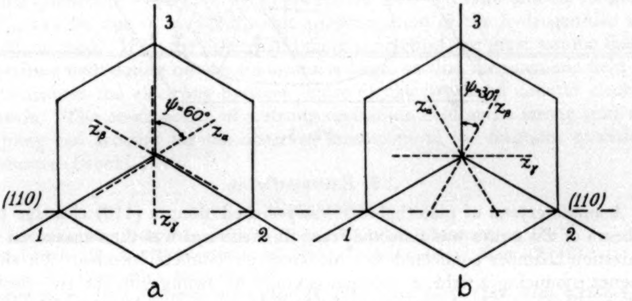


Fig. 4.

Case 1. According to fig. 4(*a*), in domains with α and β orientation the z-axis forms an angle $\psi = 60°$ with the normal to plane (110). As follows from fig. 5, $h = r \sin\theta \sin\phi \sin\psi + r \cos\theta \cos\psi$. The integration gives $f_\alpha = f_\beta = 0{\cdot}365$. The domains of γ-orientation are normal to plane (110) and thus give a value $f_\gamma = 0{\cdot}696$, which has already been calculated for case (*a*). The mean of the three values is $f_{3d} = 0{\cdot}475$.

Case 2. Domains of α and β orientation (see fig. 4 (*b*)) form an angle $\psi = 30°$ with the normal to plane (110). Using the same formulae as above we get $f_\alpha = f_\beta = 0{\cdot}588$. The domains of γ-orientation are parallel to plane (110), and thus give $f = 0{\cdot}371$ (case (*b*)).

The mean of the three values is $f_{3d}=0{\cdot}516$. In the following table all results are collected. We notice that the largest difference in intensity occurs between cases (*a*) and (*b*). It is thus established that the best way to observe changes in the intensity of x-rays reflected on plane (110), is to apply a magnetic field and to bring it first perpendicular to (110) and then parallel to it. The calculated change of S^2 amounts to 2·5%.

	f_{3d}	F	S^2
(*a*) Field perpendicular	0·696	16·66	2653
(*b*) Field parallel	0·371	16·33	2587
(*c*) No field case 1	0·475	16·43	2608
case 2	0·516	16·48	2616
(*d*) Case of spherical symmetry (Hartree)	0·44	16·40	2601

In this table we have also included case (*d*), which refers to the case of spherically symmetrical Fe atoms. The values $f_{3d}=0{\cdot}44$ and $F=16{\cdot}40$ are computed from Hartree's (1933) tables. This case would correspond to a paramagnetic pyrrhotite crystal.

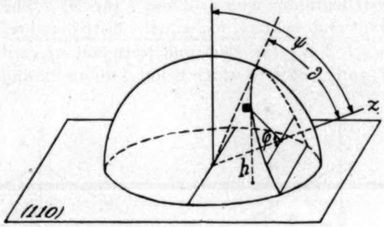


Fig. 5.

§ 3. EXPERIMENTAL

A single crystal of plate habitus was polished along a (110) cleavage plane. A beam of Cu x-rays was reflected from this face and was then measured in an ionization chamber connected to a bridge of electrometer triodes. An electromagnet producing a field of 7500 gauss could be swung into the two positions indicated in fig. 1. In the first position the field was perpendicular to plane (110). In the second it was parallel to this plane while being at the same time perpendicular to the plane of the incoming and reflected beams. The experimental arrangement has been described in detail by Theodossiou (1950).

As the changes in intensity were found to be very small, several precautions had to be taken in order to avoid systematic errors. Radiation scattered into the ionization chamber from surrounding air depends on screening by the mass of the magnet and therefore on its position; for this reason every measurement had to be compared with another with the magnet in the same position but without an exciting current. Hysteresis effects were eliminated after each measurement by a small demagnetizing coil. A strong alternating current producing a field of 90 gauss—equal to the coercive field—was passed through the coil for a very short time and then was gradually decreased. The tension and current of the x-ray tube was held as constant as possible; also the effect of the stray magnetic field was eliminated. In this way the mean error of each measurement was approximately 1%. A long series of alternate measurements gave the result that the intensity in case (*a*) was larger than that in case (*b*) by $\Delta I = 0{\cdot}3\% \pm 0{\cdot}15\%$.

§ 4. Discussion of the Results

In § 2 we calculated the change of intensity of an x-ray beam reflected on a (110) plane of pyrrhotite when the direction of the magnetic field is changed. This calculation was based on the assumption that in all Fe atoms a single 3d electron is oriented under the influence of the external field. The change should amount to 2·5%. This also holds if pyrrhotite is regarded as a ferrimagnetic (i.e. as consisting of two sublattices with antiparallel moments) as the density distribution of a 3d electron undergoes no change if the axis is turned through 180°. If on the other hand we assume that pyrrhotite is a ferromagnetic in which only every third Fe atom has a magnetic moment (§ 1) the value 2·5% diminishes to 0·83%.

As described in § 3 the effect of an external magnetic field was found to be 0·3%; the statistical error was ± 0·15%. The error is of the same order of magnitude as the measured quantity so that the existence of an effect is questionable. The effect could be further influenced by extinction. Magnetostriction can produce changes in extinction that could either enhance or lessen the intensity of the reflected x-rays.

If an effect exists, this could be accepted as a sign that the orbital moments are not quenched. The difference between the observed value and the computed, 2·5%, can be due to an insufficient approximation of the hydrogen-like wave functions used. If on the other hand there is no effect, we must assume that the crystalline field acting on the Fe atoms is large, so that the magnetic field only re-distributes the electrons between states of clockwise and counter clockwise rotation. The co-existence of a strong crystalline field and a strong spin–orbit coupling can account for the observed anisotropy in the magnetic qualities of pyrrhotite (Brooks 1940).

References

Alexopoulos, K., 1942, *Festschrift N. Kreticos* (Athens : Meissner), p. 224.
Alexopoulos, K., and Theodossiou, A., 1950, *Praktika Athens Acad.*, **25**, 419.
Brooks, H., 1940, *Phys. Rev.*, **58**, 909.
Coeterier, F., 1933, *Naturwissenschaften*, **21**, 251; 1935, *Helv. Phys. Acta*, **8**, 522.
Compton, A. H., 1921, *J. Franklin Inst.*, **192**, 145.
Compton, A. H., and Rognley, O., 1920, *Phys. Rev.*, **16**, 464.
Hartree, D. R., 1933, *Proc. Roy. Soc.* A, **141**, 282.
Hirone, T., and Tsuya, N., 1951, *Phys. Rev.*, **83**, 1063.
Inglis, D. R., 1934, *Phys. Rev.*, **45**, 118.
Néel, L., 1952, *Proc. Phys. Soc.* A, **65**, 869.
Pauling, L., and Sherman, J., 1932, *Z. Kristallogr.*, **81**, 1.
Pauling, L., and Wilson, B., 1935, *Introduction to Quantum Mechanics* (New York: McGraw-Hill), p. 132.
Stephens, R. E., 1938, *Univ. Pittsburgh Bull.*, **35**, 377.
Theodossiou, A., 1950, *Thesis*, University of Athens.
Weiss, P., and Forrer, R., 1929, *Ann. Phys., Paris*, **12**, 279.

The Velocity of Sound in Air at Low Pressures

By D. E. CARO and L. H. MARTIN
Physics Department, Melbourne University, Australia

MS. received 4th *May* 1953

Abstract. Measurements of the velocity of sound in tubes have been made in air over a range of pressures extending down to 5 mm Hg at frequencies of 250 c/s and 1000 c/s. The method employed was to measure the time of transit of a sinusoidal pulse between two condenser microphones separated by a distance of approximately 4 metres. After tube and gas corrections were made it was found that the velocity was independent of pressure between 760 mm Hg and 5 mm Hg. The mean velocity at 20°c obtained is $343{\cdot}40 \pm 0{\cdot}02$ m sec^{-1} which is in good agreement with the generally accepted value $343{\cdot}42 \pm 0{\cdot}07$ m sec^{-1} obtained by Hardy *et al.* for dry air free from carbon dioxide. No evidence was found of the increase in velocity at low pressures observed by Abbey and Barlow and by Maulard.

§ 1. Introduction

Some recent measurements of the velocity of sound in tubes, made throughout a range of pressures lower than atmospheric, have indicated that, while the velocity is independent of pressure down to pressures in the neighbourhood of 15 cm Hg, a significant increase in velocity occurs as the pressure is further diminished. In each case the measurements were made in metal tubes with audio-frequency sound. For example, in the case of Abbey and Barlow (1948) an increase in the tube velocity of 3·5 m sec^{-1} was observed when the pressure was reduced from 15 cm Hg to 0·5 cm Hg, while Maulard (1949) reported a 4% increase for the pressure range 15 cm Hg to 4 cm Hg. Since the total theoretical tube correction leads to decreasing velocities as the pressure is reduced, these results can only be explained if there is an increase in the free space velocity at low pressures. Maulard made measurements with different sound intensities and showed that the velocity increase was not due to high signal amplitudes.

Since no acceptable explanation of these observations could be found and since their validity is of meteorological significance, we have made a new determination of the velocity in air in a range of pressures extending from 760 mm Hg to 5 mm Hg. The method employed was to measure the time of transit of a sinusoidal pulse between two condenser microphones separated by a distance of approximately 4 metres. It was designed to differ significantly from the methods used by Abbey and Barlow and by Maulard, and possessed greater precision.

§ 2. Experimental Method

The microphones were mounted in the wall of a tube in the ends of which were 6-inch moving-coil loudspeakers specially constructed to operate at low pressures. The loudspeakers were rubber mounted and tests made with the gas at a pressure of 10^{-5} mm Hg showed that transmission along the metal wall of the tube was less than the microphone noise level. A pulsed sine wave was transmitted first by one loudspeaker and then by the other, the two transit times being averaged to give the true time of travel between the microphones, free of errors due to delays in the microphone circuits.

Figure 1 shows a block diagram of the whole apparatus. Consider first the production of the pulsed sine wave for the speakers. In the timing unit either a 250 c/s or 1000 c/s sine wave is produced and this is locked to a 1000 c/s standard frequency signal which is constant to 1 part in 10^6. The 250 c/s (or 1000 c/s) sine wave is passed both to the speaker gate and to the pulser, the latter producing one pulse for each cycle of the applied signals. The output from the pulser passes to a circuit which selects one single pulse whenever a firing pulse is sent to it.

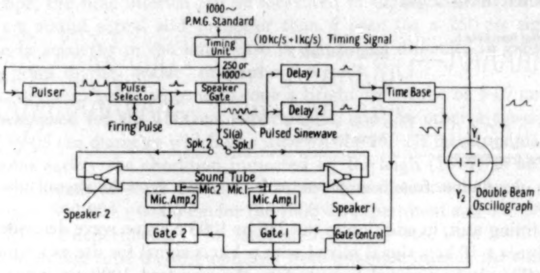

1000~
P.M.G. Standard
Timing Unit
(10kc/s + 1kc/s) Timing Signal
250 or 1000~
Pulser
Pulse Selector
Speaker Gate
Firing Pulse
Delay 1
Delay 2
Time Base
Pulsed Sinewave
S1(a)
Spk. 2
Spk. 1
Sound Tube
Mic. 2
Mic. 1
Mic. Amp. 2
Mic. Amp. 1
Speaker 2
Speaker 1
Gate 2
Gate 1
Gate Control
Y1
Y2
Double Beam Oscillograph

Fig. 1. Block diagram of apparatus.

Normally the firing pulse is produced by a push button but for testing a 10 c/s multivibrator is used. The pulse selector sends one single pulse, accurately phased with the 250 c/s (or 1000 c/s) sine wave, to the speaker gate and the phasing is adjusted so that the gate opens at the moment the sine wave has zero amplitude. The speaker gate, opened by the selected pulse, closes automatically after the desired number of sine wave cycles has passed. For 1000 c/s operation a train of 17 cycles was used, while for 250 c/s operation five cycles were transmitted.

The output from the speaker gate is connected to either speaker 1 or speaker 2, and the signal propagated down the tube is detected in turn by the two microphones. These are located three metres from the speakers so that the pulse reflected from the far end does not arrive back at either microphone until the whole of the forward moving pulse has passed.

The microphones which had diaphragms of 1·5 cm diameter, were designed for a sensitivity of 5 millivolt per microbar. The signals produced in them are amplified first by head amplifiers and then by a tuned amplifier with a *Q* of 5. This improves the signal-to-noise ratio for observations at low pressures. The use of tuned circuits has two objections. It is very easy to produce a phase shift through the amplifier, and the output signal does not rise to full amplitude instantaneously. The first defect is eliminated by making measurements of the transit times between the microphones with pulses travelling alternately in each direction. The second objection is overcome by measuring at the tenth cycle of the sine wave pulse for 1000 c/s operation and the fourth for 250 c/s operation. At these points the effects of transients have vanished.

The outputs of the microphone amplifiers pass through gates and are both applied to the Y_2 deflector plates of a double beam oscillograph. The gates are arranged so that gate 1 is open (that is passing signals) and gate 2 closed when microphone 1 is receiving a signal and conversely when microphone 2 is operating. Thus the signal presented at the Y_2 plates consist of both microphone signals one after the other. Suppose speaker 1 is transmitting so that microphone 1 receives the first signal. The resulting signal at the Y_2 plates is as shown in fig. 2(*a*). It will be noted that the output from gate 2 is at a lower mean level than from gate 1. If the sound signal is being transmitted from speaker 2, microphone 2 receives the first signal and the gate timing is reversed to give the result shown in fig. 2 (*b*).

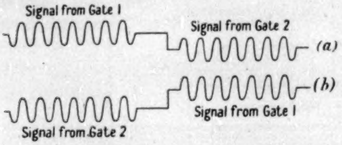


Fig. 2. Output from microphone gate.

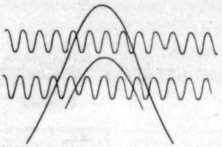


Fig. 3. Cathode-ray tube display.

The timing unit, in addition to the 250 or 1000 c/s sine wave described above, also produces a 10 kc/s signal mixed with a 1 kc/s signal for use as a time marker system. The 10 kc/s signal is locked to the standard 1000 c/s input and can consequently be treated as a standard. The 1 kc/s signal is in the form of a pulse which marks each tenth cycle of the 10 kc/s signal to aid in the counting of cycles. This timing waveform is applied to the Y_1 plates of the cathode-ray tube and it is arranged that the signal is displaced vertically at the same instant and in the same direction as the displacement of the microphone signals. In consequence, if a conventional time base was employed, the two traces would appear parallel and the time interval between say the tenth cycle at microphone 1 and the corresponding cycle at microphone 2 could be estimated by counting the number of 1 kc/s markers and the number and fraction of 10 kc/s sine waves between the two microphone signals.

Such a method would not be very precise since the time base would need to be at least as long as the transit time between microphones. Because of this, only rough time measurements are made this way (to the nearest millisecond) and accurate measurements are made with a highly expanded time base (writing speed 8 cm msec^{-1}). This time base is fired twice, once just before the peak of the cycle to be measured arrives at the microphone receiving the signal first, and the other just before the corresponding peak arrives at the other microphone. Because of the vertical displacement of the two microphone signals and the corresponding change of level of the timing signal, four traces are displayed on the screen as is shown in fig. 3. The signal from microphone 1 is always on the top of the pair of Y_2 traces and that from microphone 2 on the bottom. Thus if speaker 1 is transmitting, the top trace is produced first but if speaker 2 is transmitting, the lower trace is the earlier.

The two time-base sweeps are generated by a single time-base circuit receiving two trigger pulses from delays 1 and 2 (see fig. 1). Both delays receive their input from the pulse selector and they are adjusted so that the two time-bases start just before the peaks of the microphone signals to be measured. The pulse from delay 2, that is the later pulse, is also used to control the switching of the microphone gates.

The transit time of the signal between the microphones is measured by photographing the cathode-ray tube display. The number of milliseconds between the 1 kc/s timing markers after the first microphone signal and before the second is known from measurements with a single low-expansion time-base. The number of $\frac{1}{10}$-milliseconds is found by counting the 10 kc/s cycles forward to the marker from the first signal and back to the marker from the second. By estimating the residual fractions of a 10 kc/s cycle with the aid of a travelling microscope, the time interval can be measured to slightly better than 1 μsec for a 1000 c/s sound signal and to better than 4 μsec for a 250 c/s signal. The decrease in accuracy in the latter case is due to the difficulty in estimating the turning point of the 'wider' microphone signals.

Two different tubes were used, one a bright steel tube of 5·07 cm diameter which was used for the 1000 c/s experiments, and the other a tin-dipped steel tube of 19·05 cm diameter which was used for the 250 c/s measurements. These dimensions satisfy the condition indicated by Rayleigh (1894) as necessary for the establishment of plane waves in a tube. The distances between the microphones were 399·804 ± 0·006 cm for the 1000 c/s experiment and 395·68 ± 0·10 cm for the 250 c/s experiment. The same microphones were used with both tubes. For the narrower tube the microphones were mounted outside the tube wall and acoustically coupled to the gas by a $\frac{1}{4}$ in. hole through the tube wall. For the large tube, the microphones were mounted with their diaphragms flush with the inner tube wall.

The tube was pumped overnight at a pressure of 10^{-5} mm Hg and filled with dry air, free from carbon dioxide. The leak rate was such that during the course of an experiment the increment of pressure was never greater than 0·02%. Since most of the leak would be air, the resultant impurity of the air in the tube must have been very much less than 0·02%.

The air pressure in the tube was measured with a U-tube manometer down to 5 cm Hg and with a McLeod gauge from there downwards. An accuracy to about 1% was maintained in all pressure measurements. The gas temperature was measured with two $\frac{1}{10}$-degree mercury thermometers inserted through O ring seals near the ends of the main tube. Corrections were made for the effect of pressure variations on the thermometer bulbs. A third thermometer measured the tube wall temperature near the centre of the tube. The temperature measurements were accurate to 0·02°c. For the 1000 c/s experiment, the range of temperatures in which measurements were made was 20·7°c to 23·1°c while for the 250 c/s experiment it was 21·6°c to 25·5°c. All velocities were corrected to 20°c.

§ 3. CORRECTIONS

The corrections appropriate to velocity measurements in tubes have been considered in detail by Henry (1931). The velocities observed were corrected first to 20°c, then for variations of gas imperfection with pressure, attenuation and wall flexure. The velocities after correction were plotted against the reciprocals of the square roots of the pressures, since the form of the Helmholtz–Kirchhoff (1868) equation for the velocity of sound in a tube predicts a linear relation between these quantities.

Temperature Correction

Besides the gas density correction, allowance was made for the expansion of the steel tube. The latter was included in the reduction of the data although this correction lies just outside the final estimate of precision.

$$V_{20} = V_\theta[1+\alpha(\theta-20)]\left[\frac{293\cdot16}{273\cdot16+\theta}\right]^{1/2}$$

where V_θ is the measured velocity at θ°c and α is the coefficient of expansion of steel.

A small correction was made at this stage for the temperature variable quantities in the Helmholtz–Kirchhoff equation. According to this, the velocity of sound V_T in a tube of radius R is given by

$$V_T = V[1-\beta_1/2R(\pi n)^{1/2}]$$

where

$$\beta_1 = (\eta/\rho)^{1/2} + (\gamma-1)[k/\gamma\rho C_v]^{1/2}$$

or using Eucken's (1913) well-known relation $k = \eta C_v(9\gamma-5)/4$

$$\beta_1 = (\eta/\rho)^{1/2}\{1+\tfrac{1}{2}(\gamma-1)[(9\gamma-5)/\gamma]^{1/2}\}.$$

Here n is the frequency of the sound and V the velocity for free space; η is the viscosity, ρ the density, and γ the ratio of specific heats C_p/C_v of the gas. The correction for the variation of the viscosity of air with temperature was made with the aid of Sutherland's (1893) formula. The temperature correction to R is negligible.

Gas Imperfection Correction

This correction has been discussed at length by Hardy, Telfair and Pielemeier (1942). To the precision of the present measurements the correction for gas imperfection can be made by multiplying the velocity measured at a pressure p by $[1+3\cdot8\times10^{-7}(760-p)]$ where p is measured in mm Hg. The numerical factor was derived for 20°c from the coefficients of the virial used by Hardy *et al.*

Attenuation

On the basis of equation (40) in the paper by Henry, the correction for attenuation appropriate to the present measurements is given by the multiplying factor $\{1+\tfrac{1}{8}[(V/x\pi n)\ln A]^2\}$ where x is the distance between the microphones and A is the attenuation measured by a_1/a_2 where these are the amplitudes at the first and second microphones respectively. Experimental values for the attenuation are:

$$A = 1+16p^{-0\cdot83} \qquad (1000\ \mathrm{c/s})$$

$$A = 1+0\cdot76p^{-0\cdot68} \qquad (250\ \mathrm{c/s}).$$

Wall Flexure

The effect of the yielding of the wall of the tube has been considered by Lamb (1925). He finds that,

$$\frac{\text{velocity in actual tube}}{\text{velocity in rigid tube}} = 1-(\gamma pR/Et)$$

where E is Young's modulus for the material of the tube, and t the wall thickness, 0·32 cm and 0·63 cm for the narrow and wide tube respectively. In the present experiment, $\gamma pR/Et$ is given by $1{\cdot}49 \times 10^{-8}\, p$ for the wide tube and $7{\cdot}55 \times 10^{-9}\, p$ for the narrow tube, where p is in mm Hg.

Accommodation Coefficient Correction

Figure 4 shows corrected values of V_T at 20°c plotted against the reciprocals of the square roots of the pressures for each frequency. A least-squares fit was made for the velocity V_T in the tube as a function of the pressure p. In the case of 1000 c/s for example,

$$V_T = (343{\cdot}39 \pm 0{\cdot}04) - (17{\cdot}2 \pm 0{\cdot}45)p^{-1/2} + (3{\cdot}34 \pm 0{\cdot}97)p^{-1} \quad \ldots\ldots(1)$$

where V_T is measured in m sec^{-1} and p in mm Hg.

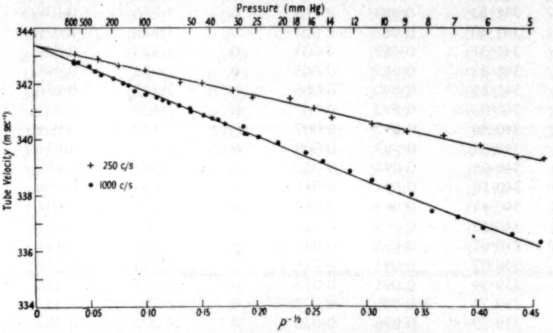


Fig. 4. Tube velocity as a function of pressure.

The slight departures from linearity observed at the lowest pressures in this, and also the 250 c/s experiment, can be accounted for completely by the accommodation coefficient correction derived by Henry.

Kirchhoff's equation modified to correct for slip between the gas and wall of the tube may be written

$$V_T = V[1 - (\beta_1 + \beta_2)/2R(\pi n)^{1/2}].$$

Here

$$\beta_2 = -\frac{\pi\eta(2/f-1)[2n/\rho]^{1/2}}{p^{1/2}[1+\pi(2/f-1)(2n\eta/p)^{1/2}]}$$

where f is the accommodation coefficient.

Since ρ is proportional to the pressure, β_2 is in the main inversely proportional to p. The bracket in the denominator is approximately unity.

The coefficient of p^{-1} in eqn. (1) leads to a value of $f = 0{\cdot}34^{+0{\cdot}11}_{-0{\cdot}07}$. This value of f is consistent with the data for the 250 c/s experiments, but the accuracy of these was insufficient to warrant a separate derivation of f for the wide tube. The value of f is not inconsistent with values generally observed for surfaces carrying an adsorbed layer of gas.

§ 4. RESULTS

The measured tube velocities corrected to 20°c, together with the relevant separate corrections, are set out in table 1 (1000 c/s) and table 2 (250 c/s). In the tables the pressure p is in mm Hg, the sound velocities and corrections are in m sec^{-1}. Column A gives the correction for gas imperfection, column B the correction for attenuation, and column C the correction for wall flexure. Column D gives the tube correction derived from the coefficient K of $p^{-1/2}$ obtained in the least squares analysis: see eqn. (1). The experimental value of $p^{1/2}\beta_1$ is given by $2RK(\pi n)^{1/2}/V$, where V is the intercept on the velocity axis. The experimental values of $p^{1/2}\beta_1$ derived from this expression are 14·23 ± 0·37 (1000 c/s), 14·7 ± 2·7 (250 c/s). The theoretical value of $p^{1/2}\beta_1$ is given by

$$(\eta p_0/\rho_0)^{1/2}\{1+\tfrac{1}{2}(\gamma-1)[(9\gamma-5)/\gamma]^{1/2}\}$$

Table 1. 1000 c/s Results

p	V_T (20°)	A	B	C	D	E	V (20°)
768	342·80	−0·001	0	0·002	0·621	−0·004	343·42
760	342·84	0	0	0·002	0·624	−0·004	343·46
624	342·76	+0·018	0	0·002	0·688	−0·005	343·46
401	342·58	0·047	0	0·001	0·859	−0·008	343·48
331	342·40	0·056	0	0·001	0·846	−0·010	343·29
279	342·26	0·062	0·001	0·001	1·031	−0·012	343·34
199	342·10	0·073	0·001	0	1·220	−0·016	343·38
156	341·94	0·078	0·002	0	1·379	−0·021	343·38
139	341·86	0·081	0·002	0	1·459	−0·024	343·38
122	341·64	0·083	0·002	0	1·557	−0·026	343·26
100·5	341·52	0·084	0·003	0	1·716	−0·032	343·29
83·7	341·40	0·088	0·004	0	1·879	−0·039	343·33
76·0	341·31	0·089	0·004	0	1·975	−0·043	343·34
69·3	341·22	0·089	0·005	0	2·074	−0·047	343·34
59·2	341·11	0·090	0·006	0	2·235	−0·054	343·39
51·1	341·02	0·092	0·007	0	2·408	−0·063	343·46
50·6	340·88	0·092	0·007	0	2·420	−0·064	343·34
44·0	340·75	0·093	0·009	0	2·597	−0·074	343·37
39·5	340·66	0·094	0·010	0	2·739	−0·082	343·42
36·8	340·58	0·094	0·011	0	2·836	−0·086	343·44
34·3	340·41	0·094	0·012	0	2·941	−0·094	343·36
31·6	340·33	0·094	0·013	0	3·061	−0·101	343·40
28·1	340·07	0·095	0·015	0	3·244	−0·113	343·31
26·8	340·03	0·095	0·016	0	3·323	−0·119	343·35
24·7	339·94	0·095	0·018	0	3·459	−0·128	343·38
20·8	339·74	0·096	0·022	0	3·770	−0·152	343·48
16·8	339·39	0·096	0·028	0	4·200	−0·187	343·53
14·9	339·09	0·096	0·031	0	4·462	−0·210	343·47
12·5	338·74	0·096	0·037	0	4·877	−0·250	343·50
10·9	338·42	0·096	0·042	0	5·224	−0·286	343·50
9·93	338·15	0·097	0·047	0	5·486	−0·314	343·47
8·68	337·88	0·097	0·054	0	5·840	−0·354	343·52
7·81	337·28	0·097	0·057	0	6·157	−0·391	343·20
6·88	337·06	0·097	0·064	0	6·562	−0·443	343·34
6·15	336·67	0·097	0·071	0	6·938	−0·491	343·29
5·39	336·45	0·097	0·080	0	7·412	−0·557	343·48
4·80	336·15	0·097	0·087	0	7·854	−0·623	343·57

Mean V=343·40 ± 0·01.

For explanation of column headings see text.

where $p_0 = 760$ mm Hg, $\rho_0 = 1{\cdot}2044 \times 10^{-3}$ g cm^{-3} (the density of air at 760 mm Hg and 20°c), $\eta = 1{\cdot}815 \times 10^{-4}$ poise. The theoretical value of $p^{1/2}\beta_1$ is 15·69. At 760 mm Hg and 20°c $\beta_1 = 0{\cdot}56_9$.

The experimental values of β_1 are thus $(90{\cdot}7 \pm 2{\cdot}7)\%$ and $(93{\cdot}5 \pm 17)\%$ respectively of the theoretical value. These values are in accord with those found by Kaye and Sherratt (1933) and Norton (1935) at atmospheric pressure.

Column E gives the correction for gas slip at the tube wall derived from the coefficient of p^{-1} in eqn. (1).

Table 2. 250 c/s Results

p	V_T(20°)	A	B	C	D	E	V(20°)
720	343·05	0·005	0	0·004	0·353	−0·001	343·41
305	342·85	0·059	0	0·002	0·541	−0·003	343·45
172	342·56	0·077	0	0·001	0·721	−0·005	343·35
100	342·48	0·086	0	0	0·945	−0·008	343·50
57·9	341·94	0·091	0	0	1·240	−0·015	343·26
44·0	341·99	0·092	0	0	1·425	−0·020	343·49
31·3	341·64	0·094	0·003	0	1·689	−0·028	343·40
26·2	341·38	0·095	0·004	0	1·851	−0·033	343·30
18·9	341·37	0·096	0·005	0	2·176	−0·046	343·60
15·9	340·98	0·096	0·006	0	2·373	−0·054	343·40
14·0	340·57	0·096	0·008	0	2·524	−0·061	343·14
10·9	340·46	0·097	0·012	0	2·867	−0·079	343·36
8·92	340·18	0·097	0·013	0	2·163	−0·095	343·36
7·39	340·06	0·097	0·017	0	3·476	−0·114	343·54
6·21	339·68	0·097	0·021	0	3·791	−0·135	343·45
5·30	339·29	0·097	0·026	0	4·104	−0·158	343·36
4·75	339·24	0·097	0·029	0	4·334	−0·176	343·52

Mean $V = 343{\cdot}41 \pm 0{\cdot}03$.

For explanation of column headings see text.

The last column gives the free space velocities at 20°c for different pressures. These are the sums of $V_T(20°)$ and the corrections A to E. It will be seen that over the range of pressures $V(20°)$ is constant to better than 1 part in 10 000 for both frequencies. The weighted mean velocity at 20°c from both experiments is $343{\cdot}40 \pm 0{\cdot}02$ m sec^{-1}, which is in good agreement with the generally accepted value $343{\cdot}42 \pm 0{\cdot}07$ m sec^{-1} (Hardy *et al.* 1942) for dry air, free from carbon dioxide.

§ 5. DISCUSSION

The experiments at both frequencies are not only consistent within themselves but also yield values of the free space velocity which are mutually consistent. It was pointed out earlier, however, that in both series of measurements the experimental value of the coefficient β_1 in the Helmholtz–Kirchhoff expression was nearly 10% less than the theoretical value. It might be argued that the present experiments represent evidence that the free space velocity increases as the pressure decreases by an amount which is nearly 10% of the corrections shown in column D. Such an effect, even if it were true, is of a different order of magnitude from the observations of Abbey and Barlow and of Maulard. In the experiments of the former the tube velocity itself increased from 342·9 m sec^{-1} at 15 cm Hg to 346 m sec^{-1} at 0·5 cm Hg. This corresponds to an increase in the free space velocity of 8·6 m sec^{-1}, ten times the increment which our measurements would show if the whole Helmholtz–Kirchhoff correction is applied. The difference is much greater for Maulard's results.

In fact there is good evidence in the earlier experiments of Kaye and Sherratt (1933), Norton (1935) and Lawley (1952) to support our observation that the theoretical coefficient β_1 is nearly 10% too great. Kaye and Sherratt found for smooth tubes of glass and copper of several diameters and for frequencies ranging from 790·8 c/s to 7908 c/s that the coefficient β_1 for air at atmospheric pressure was 91% of the theoretical value. Norton used Bakelite tubes from $\frac{1}{2}$ in. down to $\frac{3}{16}$ in. and frequencies from approximately 10 to 80 kc/s, and found an experimental value of β_1 which was 87% of the theoretical value given by Kirchhoff. The experimental coefficient in our experiment was (90·7 ± 2·3)% at 1000 c/s, and (93·5 ± 17)% at 250 c/s, of the theoretical value. In view of the satisfactory agreement between values of β_1 obtained under widely differing experimental conditions we feel justified in applying the modified correction through the range of pressures in our experiment. The theoretical expression for β_1 contains the thermal conductivity of the gas as well as the coefficient of viscosity. It has become customary to use Eucken's relation $k=\eta C_v(9\gamma-5)/4$ to obtain an expression for β_1 which contains explicitly only the kinematic viscosity and γ, the ratio of the specific heats. If we use the well-known relation of kinetic theory relating thermal conductivity to viscosity, $k=\epsilon\eta C_v$, experiment (Kannuluik and Carman 1951) gives 1·96 for ϵ at 20°C, while Eucken's expression gives 1·90. This has the effect of making our values of β_1 0·5% too low, but since there is some uncertainty in the last figure for ϵ and since this uncertainty in the tube correction lies outside our experimental error, we have used a value of β_1 resulting from Eucken's formula for the corrections in column D.

These experiments give no support for the observations of Abbey and Barlow and of Maulard. Within the limits of the accuracy of our measurements, the velocity of sound in free space is independent of the pressure down to 5 mm Hg.

ACKNOWLEDGMENT

We wish to express gratitude to Messrs. Rola (Aust.) for their kind co-operation in assembling special loudspeakers for operation at low pressures.

REFERENCES

ABBEY, R. L., and BARLOW, G. E., 1948, *Aust. J. Sci. Res.* A, **1,** 175.
EUCKEN, A., 1913, *Phys. Z.*, **14,** 324.
HARDY, H. C., TELFAIR D., and PIELEMEIER, W. H., 1942, *J. Acoust. Sci. Amer.*, **13,** 226.
HENRY, P. S. H., 1931, *Proc. Phys. Soc.*, **43,** 340.
KANNULUIK, W. G., and CARMAN, E. H., 1951, *Aust. J. Sci. Res.* A, **4,** 305.
KAYE, G. W. C., and SHERRATT, G. G., 1933, *Proc. Roy. Soc.* A, **141,** 123.
KIRCHHOFF, G., 1868, *Ann. Phys. Chem.* (Poggendorff), **134,** 177.
LAMB, H., 1925, *Dynamical Theory of Sound* (London : Arnold), p. 177.
LAWLEY, L. E., 1952, *Proc. Phys. Soc.* B, **65,** 181.
MAULARD, J., 1949, *C. R. Acad. Sci., Paris*, **229,** 1.
NORTON, G. A., 1935, *J. Acoust. Soc. Amer.*, **7,** 16.
RAYLEIGH, Lord, 1894, *Theory of Sound*, Vol. 2 (London : Macmillan), p. 161.
SUTHERLAND, W., 1893, *Phil. Mag.*, **36,** 507.

Experiments on the Propagation of Plane Sound Waves in Tubes

I: The Adiabatic Region

By D. E. WESTON*

II: The Transition Region

By D. E. WESTON* and I. D. CAMPBELL†
Physics Department, Imperial College, London

Communicated by R. W. B. Stephens; *MS. received 26th March* 1953

Abstract. I. An experimental investigation using an acoustic interferometer with a magnetostriction source is described. Results on attenuation and velocity in tubes of radius 0·013 cm and upwards, from 10 to 20 kc/s, and for a variety of tube materials and gases are given. These confirm modified Kirchhoff's formulae.

II. The velocity and attenuation of plane sound waves in tubes of radius 0·02 cm have been measured in the transition region between adiabatic and isothermal flow.

I. THE ADIABATIC REGION

§ 1. Introduction

Apart from its intrinsic interest a knowledge of the attenuation and velocity of sound is useful in many problems. In measuring the ratio of the specific heats of gases by the velocity of sound method it is necessary to calculate the velocity in a 'free' from that in a 'confined' gaseous medium. Tube attenuation is important in impedance tube measurements, instrument lines and porous absorbing media. From the time of Chladni there have been many measurements of sound velocity in tubes, and there has been much work on attenuation since that of Regnault (1868). Most of the measurements fall in the 'wide' tube region, but a few extend into the 'narrow' tube (see fig. 1 of Weston (1953), to be referred to as I). Earlier work mainly clashed with Kirchhoff's formulae, so that their validity was much questioned, but modern experiments tend to support them.

One object of the present work was to resolve the extraordinary controversy regarding Kirchhoff's formula. The predicted effects were therefore measured where they should be large, i.e. for 'wide' tubes which were however fairly small in bore.

§ 2. The Apparatus

When measuring the velocity and absorption in a very small bore tube the sound detector can be placed only at the ends. One convenient method, used in the research to be described, is to make the sound source its own detector. Pierce (1925) used this idea of the reaction on the source in the apparatus which Crandall named an acoustic interferometer. Norton (1935), May (1938, 1947), Drummond (1946) and Lawley (1952) have investigated the propagation of sound

---

* Now at the Admiralty Research Laboratory, Teddington.
† Scientific Officer of the Commonwealth of Australia, Department of Supply.

in small tubes using a magnetostriction source, in most cases a small monel needle for frequencies of the order of 100 kc/s.

In the author's ultrasonic interferometer high-Q monel rods of 0·6 cm diameter were used, these being held horizontally and clamped in a rubber washer at the centre. Two coils on bobbins were fitted on either side, shielded from one another by mumetal (see Knight (1950) for a full description).

The length of air column in the tube could be varied in two ways, by the withdrawal of a closely fitting piston of soft metal along the tube, or by moving a mercury thread in the tube. This air column in the experimental tube could be coupled to the source in two ways, by placing the tube in close proximity to the plane end of the oscillating rod, or by connecting through a half-wavelength brass tube of greater diameter, which acts as an acoustic transformer and increases the coupling to the experimental tube.

The analysis of the pressures in the tube and the reaction of the source has been undertaken by many writers, beginning with Hubbard (1931) and Hershberger (1931). Conditions in a small bore tube are in some respects simpler than those in a larger interferometer, i.e. there is no possibility of higher order acoustic modes being propagated, and the reflection surface is comparatively unimportant.

Making simple assumptions about the transmission and reflection coefficients for the point where the experimental and coupling tubes meet, it may be shown that the acoustic transformer ratio is equal to the ratio of the cross-sectional areas of these two tubes. Thus the sensitivity should be multipled by the (area ratio)^2, but in practice this comes nowhere near realization because the transformer is inefficient and introduces extra losses.

A self-excited electromechanical system was used, with the driving force on the rod held constant, and the velocity amplitude as the measured variable. The rod and tube system was placed in an aluminium box, and the temperature thermostatically controlled, since monel metal has a particularly low Curie point (Cook 1950). The system was almost airtight, so that slow streams of different gases could be passed into the apparatus, after drying and purification.

§ 3. PROCEDURE

After the conditions inside the box had become stable the reflector was slowly withdrawn. The attenuation was calculated from the equation, initially derived empirically by Pielemeier (1929):

$$m' = \frac{\ln 10}{\lambda} \frac{d \log \Delta I}{dn}.$$

Here n is the number of the resonance, and ΔI is the difference between a current minimum and the mean of the two neighbouring maxima.

The theory of the interferometer indicates that this equation is only valid for the larger values of n, and the first few readings of ΔI were therefore ignored; and it may also be shown that for moderately high attenuations there is some advantage in using the definition of ΔI given above. It was also necessary to satisfy the condition that the current measured be a linear function of the out-of-phase component of the excess gas pressure at the rod face. Hardy (1943) considers these two points, but his experimental proof of the high value of m' resulting from a violation of the second condition is open to doubt. With this method a knowledge of the absolute value of the current is unnecessary, and it also allows automatically for a slow and steady current drift. The attenuation in air could usually be determined to within ± 1%, the mean of several values being taken. The wavelength error varied from ± 0·03% to ± 0·5% in the smallest tube, the latter using a mercury thread as reflector.

An alternative method of finding m' is to measure the peak widths of the current–distance curve as the reflector is moved slowly along. It is more laborious and less accurate, but gives in addition a value for the product of the reflection coefficients at source and piston. In a particular case the two techniques gave the same value for m', and the product above was 0·75.

§ 4. RESULTS

All the results quoted are corrected to 20°C and 76 cm Hg. Preliminary measurements showed that the attenuation was not altered by varying the intensity through a 25 : 1 range, nor by the use of an imperfect reflector (reflection coefficient about 0·1). For a radius of 0·1 cm and a frequency of 10 kc/s no difference could be detected between the absorptions for dry and saturated air.

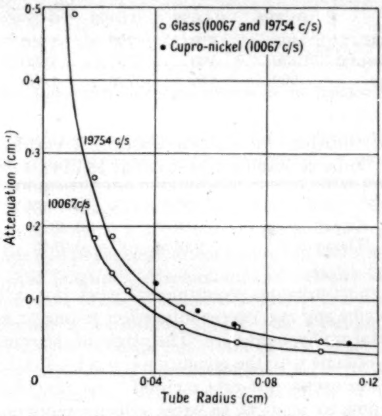


Fig. 1. Theoretical and measured attenuations in glass and cupro-nickel tubes.

Most measurements were made in air using Veridia precision bore glass tubing, which appeared to be better than ± 0·0005 cm in radius. Velocity and absorption measurements were made between 10 and 20 kc/s. Figure 1 shows the attenuations for a range of tube radii at the two extreme frequencies, which are compared with the theoretical curves calculated from eqn. (45) of I. The velocity measurements at 20 kc/s only have been chosen as a typical set. In table 1 the experimental figures for the velocity reduction (assuming a velocity in unconfined air 34 340 cm sec^{-1}) are compared with the theoretical values found from eqn. (44) of I.

Measurements were made for cupro-nickel tubes (fig. 1), and also with other tube materials, i.e. glass, stainless steel, and copper; some of these are fully reported by Weston (1950). All showed good agreement of experiment with theory as regards both velocity and absorption, except where tube imperfections were known to exist. Measurements for a number of gases in Veridia glass tubes were made at 20 kc/s, and the results for a tube radius 0·0498 cm are compared with theory in table 2.

§ 5. Discussion of Results

Attenuation and velocity measurements have been made in the region where the acoustic boundary layer is becoming comparable with the tube radius, and the results are in fair agreement with the modified Kirchhoff's formulae. Any large discrepancies could be traced to imperfections in the tube wall. Other factors affecting the validity of Kirchhoff's formulae (listed in I) may be shown to be negligible in this research—e.g. tube wall yielding is negligible—and even for the most unfavourable case slip and temperature drop effects increase the Kirchhoff value of velocity by only about 40 cm sec^{-1}.

Table 1. Velocity Reductions for Air in Veridia Glass Tubes at 19 754 c/s

The 'modified' Kirchhoff velocity reductions are those calculated from I, eqn. (44).

Tube radius (cm)		0·0996	0·0698	0·0498	0·0248	0·0186	0·0127
Velocity reduction (cm sec^{-1})	Kirchhoff	400	570	790	1560	2130	3120
	Experimental	380	590	770	1480	2010	2890
	Modified	400	570	770	1480	2020	2880

Table 2. Attenuations for Various Gases in a Veridia Glass Tube of Radius 0·0498 cm at 19 754 c/s

Gas		Air	Nitrogen	Argon
Attenuation m' (cm^{-1})	Experimental	0·0889	0·0828	0·112
	Theoretical	0·0856	0·0846	0·105

It is possible in most previous experimental work (for full bibliography see Weston 1952) to ascribe any real excess tube effect to one or more of the factors in the list, in particular wall roughness. The origin of the remarkable agreement which Vance (1932) obtains with the equantion $c' = c(1 - 0{\cdot}00128/r_w^2)$ is not clear, but it may be because he uses a dust-tube method. Kundt (1868) himself remarks that too much fine powder leads to an extra velocity reduction; the writer has obtained in this way a 20% decrease in velocity in a 0·1 cm bore glass tube, at a frequency of 20 kc/s.

It is concluded that Kirchhoff's formulae give a correct quantitative description of propagation in a 'wide' tube, provided the modifications and conditions discussed in the previous paper are borne in mind.

II. THE TRANSITION REGION

§ 6. Introduction

No satisfactory measurements for the transition region between the 'wide' and 'narrow' tube are known to the authors, although several previous experimenters have worked on the edge of the 'wide' tube region (as in part I of this paper). The researches of Simmons and Johansen (1925), Richardson (1926) and Nief (1946) lie in the transition region, but no precise results on velocity and attenuation were presented. In addition several measurements in the transition region have been made of the impedance of stacks of tubes, simulating an acoustic absorber. This paper describes single tube measurements designed to give results which can be compared with the theoretical curves of paper I.

§ 7. METHOD

Observations with a single tube of radius about 0·02 cm, containing air at normal pressure, were made at frequencies varying from 70 c/s to 8 kc/s. A tube T of this size required the source S (a dynamic pressure unit) and the detector M (a crystal microphone) to be each located at one end, cf. fig. 2. By using tube lengths up to four feet it was arranged that the attenuation was always sufficiently great to eliminate standing waves. The magnitude of the transmitted signal was measured by comparing the microphone output, after pre-amplification, with a known fraction of the voltage exciting the loudspeaker. Similarly a comparison of phase could be made using a double-beam cathode-ray oscilloscope.

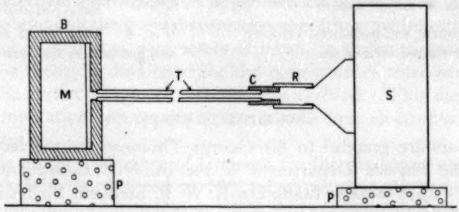


Fig. 2. Apparatus for measurements in the transition region.

These amplitude and phase comparisons were made at a number of frequencies for a number of tube lengths. Attenuation and velocity are directly calculable from the amplitude and phase changes. Basically this method of finding the propagation constants is extremely simple—for example, provided the end conditions are unchanged, it is not important to have exact impedance matching. Due to the small signal received, precautions such as electrical screening, attention to electrical noise reduction, and acoustic isolation of the loudspeaker and microphone were necessary. Sound transmission through the tube wall was prevented by a rubber connection R, and the possibility of external air-borne sound affecting the microphone was guarded against by a suitable massive brass housing B mounted on soft rubber pads P. A check on the system as a whole was effected by inserting a solid dummy connection between the tube under test and the brass connecting piece C, when it was found that there was no signal above the background noise.

§ 8. RESULTS

The results of two series of measurements are shown in figs. 3 and 4, the first with a cupro-nickel tube and the second with one of stainless steel. On the whole the attenuation results are a little higher than the calculated values, though they follow the predicted variation with frequency. The velocity measurements are in general accord with theory.

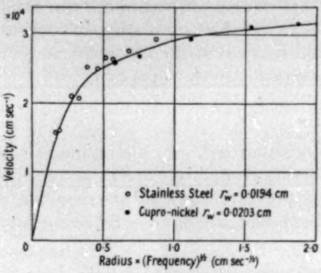


Fig. 3. Theoretical and measured velocity in the transition region.

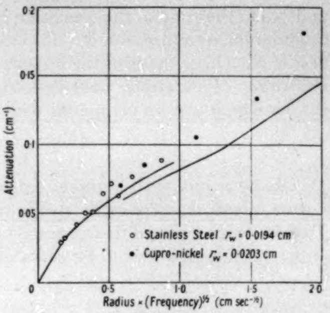


Fig. 4. Theoretical and measured attenuation in the transition region.

ACKNOWLEDGMENTS

The authors are grateful to Sir George Thomson for the facilities at their disposal in the Physics Department of the Imperial College of Science and Technology. They wish to thank Dr. R. W. B. Stephens for his supervision of the work and for valuable discussions. Finally one of them (D.E.W.) must express his thanks to Mr. J. J. Knight for the privilege of using the magnetostriction oscillator mounting designed by him.

REFERENCES

COOK, J. C., 1950, *Nuovo Cim.*, Suppl No. 2, **8,** 192.
DRUMMOND, J. E., 1946, *Thesis*, University of Otago.
HARDY, H. C., 1943, *J. Acoust. Soc. Amer.*, **15,** 91.
HERSHBERGER, W. D., 1931, *J. Acoust. Soc. Amer.*, **3,** 263.
HUBBARD, J. C., 1931, *Phys. Rev.*, **38,** 1011.
KNIGHT, J. J., 1950, *Nuovo Cim.*, Suppl. No. 2, **8,** 392.
KUNDT, A., 1868, *Ann. Phys., Lpz.*, **135,** 337.
LAWLEY, L. E., 1952, *Proc. Phys. Soc.* B, **65,** 181.
MAY, J., 1938, *Proc. Phys. Soc.*, **50,** 553; 1947, *Ibid.*, **59,** 33.
NIEF, G., 1946, *C. R. Acad. Sci., Paris*, **223,** 306.
NORTON, G. A., 1935, *J. Acoust. Soc. Amer.*, **7,** 16.
PIELEMEIER, W. H., 1929, *Phys. Rev.*, **34,** 1184.
PIERCE, G. W., 1925, *Proc. Amer. Acad. Sci.*, **60,** 271.
REGNAULT, V., 1868, *C. R. Acad. Sci., Paris*, **66,** 209.
RICHARDSON, E. G., 1926, *Proc. Roy. Soc.* A, **112,** 522.
SIMMONS, L. F. G., and JOHANSEN, F. C., 1925, *Phil. Mag.*, **50,** 553.
VANCE, C. B., 1932, *Phys. Rev.*, **39,** 737.
WESTON, D. E., 1950, *Nuovo Cim.*, Suppl. No. 2, **8,** 341; 1952, *Thesis,* University of London; 1953, *Proc. Phys. Soc.* B, **66,** 695.

Correction of Electrostatic Lenses by Departure from Rotational Symmetry

By J. C. BURFOOT*
The Cavendish Laboratory, Cambridge

Communicated by V. E. Cosslett; MS. received 5*th March* 1953,
read before the Society on 15*th May* 1953

Abstract. It is known that departures from rotational symmetry may offer correction possibilities for certain aberrations which limit the resolving power of present-day electron microscopes. The geometrical optics of such departures in electrostatic systems and the general correction conditions are investigated. The minimum number of design parameters is taken, in order to ensure a simple structure, and a theory is developed for the correction of third-order spherical aberration in the particular type of objective lens evolved. This theory is simple enough for computation, yet applies to lenses with four electrodes capable of a high magnification.

An example shows a tolerance of $\frac{1}{8}$ micron for the correcting electrode if the resolution limit is to be reduced to 10% of its rotationally symmetric value. A better example could probably bring this within engineering practice, but first-order astigmatic electrodes are also needed, and the tolerances of these are one or two factors of ten more exacting. It is therefore here that the difficulties will arise in the application of non-rotational symmetry.

§ 1. INTRODUCTION

RESOLVING powers attained in present-day electron microscopes are apparently limited by imperfect symmetry in construction or alignment (Hillier 1946), but if we assume this difficulty to have been overcome, the limit is set by the geometrical aberrations of the lenses, at present kept below 10 Å by limiting severely the objective lens apertures. Under these circumstances, the fifth-order aberrations are negligible (Seeliger 1949), though often undesirably long exposure times are needed. Reduction of the third-order aberrations would therefore improve the resolution. Third-order spherical aberration of the objective lens is by far the most important of these.

Scherzer (1936) has shown that neither the first-order chromatic aberration on the axis nor the third-order spherical aberration can ever be made zero, the demonstration being valid for static rotationally symmetric lenses free of space charge, and provided that the axial electric potential does not fall to cathode potential at some point, as it does in electron mirrors. In the search for higher resolutions, it is clearly of importance to investigate all possibilities of escape from Scherzer's theorem. Absolute correction within the limitations of the theorem is unattainable, though for completeness it should be remembered that the aberrations may nevertheless always be reduced below any prescribed limit in principle, even for fixed lens power (Rebsch 1938). It is not necessarily true that such progress within the limitations of the theorem would lead to better starting points for more general attacks. The four conditions which limit the validity of the theorem offer four possible evasions of its consequences, namely, the use of non-static (high frequency) lenses, lenses involving appreciable space charge, electron mirrors, or lenses which depart from rotational symmetry.

---

* Now at the University of Aberdeen.

All these methods present formidable difficulties (Scherzer 1947, 1949). This paper considers only the last-mentioned for electrostatic systems. The object is to investigate theoretically whether the method offers practicable possibilities for the correction of third-order spherical aberration. Chromatic aberration is a less urgent problem, and will not be considered. Reduction of the third-order spherical aberration such as to improve the resolution by a factor g might alternatively be exploited as an increase of aperture by a factor $g^{1/3}$, with unaltered resolution. Exposure times would be shortened by $g^{2/3}$, but the fifth-order aperture aberrations would also become worse by $g^{5/3}$, and so probably no longer negligible. In addition, a loss of contrast might be incurred. This alternative use will not be considered.

Under conditions of rotational symmetry, 'even-order' aberrations are absent but, in general, departures from rotational symmetry will introduce effects of all orders, including the even orders. Scherzer, however, has considered not quite general conditions, in which symmetry is retained about each of two mutually perpendicular planes, and it may easily be seen that even-order aberrations are again absent. He has given explicit expressions for the third-order spherical aberration under these conditions. The expressions contain a number of independent functions, which may be adjusted in the attempt to make this aberration zero at the chosen image plane, but they are in a form which makes it difficult to proceed logically towards a corrected system in a straightforward manner. In this paper is shown a way to carry out the analysis in a different manner which avoids this difficulty. The new expressions also display independently the first-order astigmatism and first-order distortion which accompany the departure from rotational symmetry. Further, it becomes possible to count the minimum number of degrees of freedom of design necessary for the correction.

Seeliger (1949) has partially demonstrated the possibilities of non-rotational symmetries for correction, but the complete system he proposed contains many components and makes use of an intermediate image. The expressions here developed are used to show that in principle the minimum number of design parameters can conveniently be provided using only four electrodes. No intermediate image is used, and the electron paths through the suggested lens are similar to those in the three-diaphragm unipotential system frequently used as objective lens. In a particular case corresponding to a practical lens, the exact shapes of the apertures, through which the electron beam passes, have been calculated. They are not simple, and the calculation is carried out, not as a design procedure, but in order to estimate the order of magnitude of the design tolerances. It is found that those electrodes which control the first-order astigmatism are much more critical than the electrode which specifically performs the third-order correction. It is thought that this will be found to be a principle of general validity.

§2 determines the general correction conditions; §3 suggests the particular four-electrode system; §4 shows how the axial requirements may be found, and §5 gives a numerical example. A few possibilities are suggested for further investigation.

§ 2. Ray Equations and Aberration Expressions

Electron paths through the lens are expressed by giving their lateral coordinates, cylindrical or cartesian, as functions of the longitudinal coordinate z, thus $r=r(z)$, $\theta=\theta(z)$, or $x=x(z)$, $y=y(z)$.

The governing variational principle determining the paths taken by electrons between two points A and B in an electrostatic field is that the action is a minimum along the actual path, by comparison with neighbouring lines between A and B. Or it may be written

$$\delta\int_{\mathrm{A}}^{\mathrm{B}} p\,ds=0.$$

If the zero of the electric potential is taken to be where the electrons have zero velocity, the energy relation is

$$\mathbf{e}\phi=(m-m_0)\mathbf{c}^2, \qquad \ldots\ldots(1)$$

where ϕ is the electric potential at any point, m is the mass of the electron when travelling with the momentum p appropriate to that point, and the other symbols have their usual meanings. Re-writing this relation to show p explicitly, and using it in the variational principle, this becomes

$$\delta\int_{\mathrm{A}}^{\mathrm{B}}\{\phi(1+\kappa\phi)(1+x'^2+y'^2)\}^{1/2}\,dz=0, \qquad \ldots\ldots(2)$$

where $\kappa\equiv\mathbf{e}/2m_0\mathbf{c}^2=0{\cdot}000978$ per kilovolt. Primes denote differentiation with respect to z, and x', y' are characteristics of the ray paths. The ray equations are Euler equations of (2), and in the case of rotational symmetry, $(1+x'^2+y'^2)$ becomes $(1+r'^2)$ and only one Euler equation results. For non-rotational symmetry, however, rays cannot in general remain throughout in a meridian plane $\theta=$ constant, so that there are two Euler equations, which may be written

$$\left.\begin{aligned}2\phi x''&=\left(\frac{\partial\phi}{\partial x}-x'\frac{\partial\phi}{\partial z}\right)(1+x'^2+y'^2)\left(\frac{1+2\kappa\phi}{1+\kappa\phi}\right)\\ 2\phi y''&=\left(\frac{\partial\phi}{\partial y}-y'\frac{\partial\phi}{\partial z}\right)(1+x'^2+y'^2)\left(\frac{1+2\kappa\phi}{1+\kappa\phi}\right).\end{aligned}\right\} \qquad \ldots\ldots(3)$$

The aberration expressions are conveniently given as integrals of functions of z, so that it is necessary to express $\phi(r,\theta,z)$, using now cylindrical coordinates, in terms of the axial potential $\Phi(z)$. Inserting the condition that ϕ must satisfy Laplace's equation, one finds an expression which, for the case of symmetry about $\theta=0$ and $\theta=\pi/2$, becomes (Scherzer 1947)

$$\begin{aligned}\phi=\Phi&+r^2\left[-\frac{\Phi''}{2^2}+\phi_2\cos 2\theta\right]\\ &+r^4\left[\frac{\Phi^{(4)}}{2^2.4^2}-\frac{\phi_2''}{4^2-2^2}\cos 2\theta+\phi_4\cos 4\theta\right]\\ &+\ldots\end{aligned} \qquad \ldots\ldots(4)$$

The terms independent of θ form the well-known expansion of $\phi(r,\theta,z)$ in terms of $\Phi(z)$ in the case of rotational symmetry, while $\phi_2(z)$, $\phi_4(z)$, $\phi_6(z)$, ... are independent functions characterizing the departures from rotational symmetry. Substituting (4) in (3), it is readily seen that the first degree terms in r, which will lead to the first-order lens properties, now involve $\phi_2(z)$ in addition to the rotational symmetry function $\Phi(z)$, while the third-order terms involve $\phi_2(z)$ and $\phi_4(z)$, associated respectively with $\cos 2\theta$ and $\cos 4\theta$ components of symmetry. Fifth-order terms involve $\phi_2(z)$, $\phi_4(z)$, and $\phi_6(z)$; and so on.

Reverting now to the cartesian coordinates in which (3) was expressed, the substitution of (4) gives $L(x)=F(x,y)$ where*

$$
\begin{aligned}
L(x) &\equiv x''+\left(\frac{1}{2}\frac{\Phi'}{\Phi}x'+\frac{1}{4}\frac{\Phi''}{\Phi}x-\frac{\phi_2}{\Phi}x\right)\frac{1+2\kappa\Phi}{1+\kappa\Phi}\\
F(x,y) &\equiv x'y'y''-x''y'^2+\frac{1}{4}\frac{\Phi''}{\Phi}x''(x^2+y^2)+\frac{1}{32}\frac{\Phi^{(4)}}{\Phi}x(x^2+y^2)-\frac{1}{2}\frac{\Phi''}{\Phi}xy'^2\\
&\quad-\frac{1}{4}\frac{\Phi''}{\Phi}xx'^2+\frac{1}{8}\frac{\Phi'''}{\Phi}x'(x^2+y^2)-\frac{1}{2}\frac{\Phi'}{\Phi}x'(x'^2+y'^2)+\frac{1}{4}\frac{\Phi''}{\Phi}x'yy'\\
&\quad-\frac{\phi_2}{\Phi}x''(x^2-y^2)-\frac{1}{6}\frac{\phi_2''}{\Phi}x^3+\frac{\phi_2}{\Phi}2xy'^2-\frac{\phi_2'}{2\Phi}x'(x^2-y^2)+\frac{\phi_2}{\Phi}x'^2x\\
&\quad+\frac{\phi_2}{\Phi}x'yy'+\frac{\phi_4}{\Phi}2x(x^2-3y^2). \qquad \ldots\ldots(5)
\end{aligned}
$$

In the 'twin' equation, x and y are interchanged and $\phi_2, \phi_6, \phi_{10}, \ldots$ are changed in sign. L contains the linear terms; F represents the third-order aberrations.

The classical method of solution is to solve the first-order parts $L=0$ and substitute this first approximation into the higher-order terms. The first-order image planes for the x and y coordinates of the electron path will not in general coincide, so that the imaging is astigmatic (fig. 1 and Appendix I). The x and y magnifications will also be unequal, so that first-order distortion exists. These effects of $\phi_2(z)$ are not to be confused with third-order astigmatism and distortion respectively, which are present even in rotationally symmetric systems; for example, rotationally symmetric third-order distortion will affect the size of the image of a central circle, while this first-order distortion will affect its shape, resulting in an elliptical image.

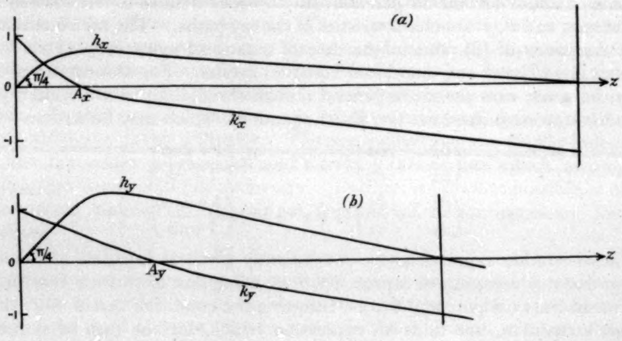


Fig. 1. Astigmatic paraxial solutions. (*a*) At the image plane for x coordinate of the rays h_x is zero. (*b*) At the image plane for y coordinate of the rays h_y is zero.

The third-order aberrations are then represented by definite integrals of functions of z, taken from object plane to image plane. They are to be made to vanish at a particular plane (i.e. for a particular value of the upper limit) by adjusting the functions $\phi_2(z)$ and $\phi_4(z)$ which occur explicitly in the integrands. In Scherzer's analysis, the integrands also contain the first-order astigmatic solutions x_α, y_α (Appendix I). These involve $\phi_2(z)$ implicitly, in such a manner that any adjustment made to $\phi_2(z)$ at any longitudinal position z causes alterations to x_α, y_α at all succeeding positions, in addition to altering the first-order conditions concerning astigmatism and distortion. Thus to design a corrected lens by logical computation, it would be necessary to undertake a series of successive alternate readjustments, *within* another such series, and alternating with renewed integrations.

---

* $F(x, y)$ contains also third-degree κ terms, and chromatic terms, which have been given elsewhere (Burfoot 1952 b) but are not required in this paper.

The method adopted in this paper avoids this difficulty by accepting a poorer first approximation, namely the non-astigmatic parts of x_α, y_α. This is done by transfering from L to F in (5) the first-order astigmatism term ϕ_2. The first-order astigmatic solutions may then be written in the form (Appendix I)

$$\left.\begin{aligned} x_\alpha &= \frac{x_\mathrm{A}}{h_\mathrm{A}}(h + M_1 + M_2 + M_3 + \dots) + x_\mathrm{O}(k + M_1{}^\mathrm{O} + M_2{}^\mathrm{O} + M_3{}^\mathrm{O} + \dots) \\ y_\alpha &= \frac{y_\mathrm{A}}{h_\mathrm{A}}(h - M_1 + M_2 - M_3 + \dots) + y_\mathrm{O}(k - M_1{}^\mathrm{O} + M_2{}^\mathrm{O} - M_3{}^\mathrm{O} + \dots) \end{aligned}\right\} \quad \dots(6)$$

where $h(z)$, $k(z)$ are special solutions of the rotationally symmetric parts, which may be determined by first-order ray-tracing methods, x_O, y_O, x_A, y_A are the values of the non-astigmatic solution at the object and at position A (fig. 1) respectively, and the M terms are successive modifications to these functions. The first-order condition for no astigmatism, with respect to any chosen image plane $z = z_\mathrm{I}$, is shown to be that the values of $x(z)$ and $y(z)$ at z_I, should be independent of x_A, y_A. This condition may be written

$$(h + M_2 + M_4 + \dots)_\mathrm{I} = 0, \qquad \dots\dots(7)$$

$$(M_1 + M_3 + M_5 \dots)_\mathrm{I} = 0. \qquad \dots\dots(8)$$

Similarly the condition for no first-order distortion is

$$(M_1' + M_3' + M_5' + \dots)_\mathrm{I} = 0. \qquad \dots\dots(9)$$

Suitable choice of the function $\phi_2(z)$ may satisfy these conditions. Finally (6) may be substituted into the integrals of (A 2) to obtain the higher-order conditions for no aberration. These conditions involve $\phi_2(z)$, which is now fixed, and $\phi_4(z)$, $\phi_6(z)$ ... which are therefore to be adjusted to achieve the desired lens correction. In particular, the ϕ_4 term of (5) enables F to be adjusted for third-order correction.

§ 3. THE FOUR-ELECTRODE LENS

The feasibility of correcting electron lenses by means of departures from rotational symmetry has been partially demonstrated by Seeliger (1949) but the complete system proposed in the case of third-order spherical aberration, involves four lenses each comprising three slits, together with three special ϕ_4 electrodes. This system uses strongly astigmatic intermediate images. With the preceding analysis, however, it has been found possible, at least in principle, to use as few as four electrodes. It is not necessary for intermediate images to exist, provided that a strongly astigmatic region occurs, in which region ϕ_4 is to operate to achieve third-order correction. ϕ_2, ϕ_4 ... may be regarded as providing a number of adjustable controls of the lens properties or degrees of freedom of design. Examination of the integrands has shown that the leading ϕ_4 term can provide only two such third-order controls which are independent, and that a minimum of four independent first-order controls must be provided in addition, using $\phi_2(z)$, if third-order correction is required. An additional ϕ_2 control will be needed if a certain image plane or the magnification is to be stipulated. If simultaneous chromatic correction were required, further controls would be needed. It has been seen that ϕ_6, ϕ_8 ... affect only fifth-order aberrations and above, and will not concern us.

The analysis can determine only the axial functions $\phi_2(z)$, $\phi_4(z)$; the realization of these functions by an electrode structure is a further, distinct, step. However, an embarrassingly large variety of functions is possible, and this choice may be limited initially by restriction to functions which correspond to reasonable electrode structures. Thus if the potential distribution of the lens of fig. 2(*a*) is taken for the rotationally symmetric parts of the lens (to be considered non-adjustable for purposes of this investigation), we may choose to consider, for representation of the non-rotationally symmetric parts, only those functions $\phi_2(z)$, $\phi_4(z)$ which consist of positive or negative 'peaks' located near the electrode apertures, since these will correspond to modifications of the aperture shapes, originally circular. If the shape of a function peak is built up by addition of several independent shapes, the ϕ_2 controls might consist of numbers representing the amount of each component shape present. However, the precise shape will be closely correlated only with the *longitudinal* sections or 'profiles' of an electrode which may be used to produce it, while it is the *cross* section of the electrode or aperture which is of more concern, being characterized by the symmetry component ($\cos 2\theta$ in the case of ϕ_2) and determined by the magnitude of the ϕ_2 effect. It is thus preferable to fix the shape of the function peaks arbitrarily, and associate one ϕ_2 control with each peak, in the form of a multiplying constant for this standard peak shape.

The requirements then suggest a lens consisting of the potential distribution of the strong unipotential lens of fig. 2(*a*) as a rotationally symmetric basis, with an astigmatism applied at the first electrode by means of a peak of the $\phi_2(z)$ function there, and partially removed by a $\phi_2(z)$ peak of opposite sign at the third electrode (fig. 2), the ϕ_4 controls for third order correction being applied in the strongly astigmatic region at the second electrode. The remaining astigmatism is removed, with respect to the final image plane, by a fourth electrode, which can be situated in the previously field-free space nearer the image region, where electrons travel rectilinearly; this choice of its situation simplifies the expressions considerably. One of the two extra controls may be associated with this electrode, by means of a multiplying constant as before, and the other by means of its longitudinal position z_C. The magnitude of the deflections produced by the fourth electrode increases with the multiplying constant, which can therefore be regarded as a measure of its 'strength'. This electrode does not provide any component of rotational symmetry.

Elliptical apertures or slits do not produce only ϕ_2 components, and the other unwanted components affect the higher order aberrations, and may make them no longer negligible even when retaining very small lens apertures*. Specially designed aperture shapes are required, which produce only rotational symmetry effects and ϕ_2 effects. However, it may be of use to visualize the first three apertures as in some respects similar to a lens with 'crossed ellipses', with the ϕ_2 multiplying constants m and n, respectively, analogous to the 'ellipticities' (fig. 2(*b*)).

---

* The lens aperture, which may be 5×10^{-3} radian, is not set by the apertures in the lens electrodes.

The condition for no astigmatism, namely that the zeros of h_x, h_y must be coincident, means that (6) must have coincident zeros for the case $x_0 = y_0 = 0$ (Appendix I). So the condition is that an electron from the object 'centre' must pass again through the axis at some position. Using (3) and (A 2) it may be shown that the familiar theorem of Lagrange–Helmholtz in light optics (the 'sine condition') holds also for electron optics,* so that magnification is equal to the ratio of the angles made with the axis in the field-free object and image regions by a track between the object and image centres, if the refractive index is the same in both regions, i.e. if the lens is unipotential. Similarly the magnification in each direction is the ratio of the appropriate projections of these angles, the classical sine condition being represented in non-rotational symmetries by two corresponding conditions. The condition for no distortion, therefore, is that an electron leaving the object centre in any particular meridian plane through the axis must travel parallel to that plane in the image region. The first-order action of the four-electrode lens on such an electron may now be described. The electron is turned out of the meridian plane as it passes through the first aperture, and turned towards that plane again as it traverses the third aperture. It will be seen later that the fourth aperture is situated near the position where the electron crosses the original meridian plane again, and this electrode turns the electron again into that plane, where it travels at a certain 'approach' angle, its eventual intersection with the axis determining the image plane. The approach of this electron to the image in a direction parallel to the original meridian plane would

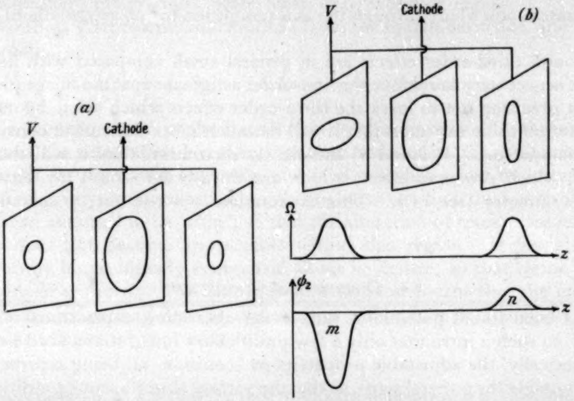


Fig. 2. (*a*) Schematic diagram of the symmetrical three-aperture unipotential lens, with central electrode at cathode potential. (*b*) The 'crossed ellipse' electrodes of the four-electrode lens, and the Ω and ϕ_2 functions.

---

* This has been shown also thermodynamically by Zworykin *et al.* (1945).

therefore ensure the absence of first-order distortion, while its approach *in* that plane ensures the absence of astigmatism also. The action on an electron leaving a more general object point is similar but less simple.

The same conclusions may be drawn from (9), which states that (8) holds, not only at z_1 ,but throughout the field-free region beyond the fourth electrode. Equation (6) then shows that an electron from the object centre travels in its original meridian plane beyond the fourth electrode (see fig. 3). This description of the first-order action confirms that the number of first-order parameters provided in this lens does not exceed the minimum possible number. For example, if it is proposed to take a fourth electrode position other than that predicted above, and to compensate this change by a change in the ' strength ' parameter, it is clearly possible to choose the latter *either* so that an electron from the object centre travels parallel to the original meridian plane (condition for no distortion) *or* so that it passes somewhere through the axis (condition for no astigmatism), but not both.

Although third-order effects are in general small compared with first-order effects, it is necessary to remove the first-order astigmatism at the image plane with sufficient precision not to mask the third-order effects which are to be corrected. It is therefore to be expected that it will be much more difficult to construct the astigmatic apertures (1, 3 and 4) sufficiently accurately, than it will the second aperture, which produces effects which are third-order simply by virtue of its type of symmetry (see (4)). This expectation is borne out by specific calculations.

§ 4. CORRECTION CONDITIONS

It has been found possible to reduce the aberration expressions which are obtained, to such a form that only a few preliminary integrations need be carried out numerically, the adjustable properties or ' controls ' all being represented by symbols outside the integral signs, so that the various simultaneous conditions may be satisfied simply by algebraic adjustment of these symbols. In this way expressions may be derived for each of the control parameters in terms of the chosen rotationally symmetric properties and object position. The procedure may be outlined as follows.

The paraxial solutions *h* and *k* in (6) are there modified (cf. (A 3)) by a series of decreasing terms M_1, M_2, ... In substituting these into the integral expressions for the third-order aberrations, modifications at least as far as M_1 must be retained, while in using (6) directly to study the first-order conditions comparable exactness will require the retention of more terms. But the successive modifications contain integrals of increasing multiplicity. Retaining only these most significant terms, the third-order non-relativistic terms which impair the imaging of the object centre (spherical aberration) are found, and the third-order correction conditions are shown (Appendix I) to take the form

$$\left.\begin{aligned} I+e+\zeta-(h/k)_1(\bar{I}+\bar{e}+\bar{\zeta})+\chi_e&=0\\ I+g-3\zeta-(h/k)_1(\bar{I}+\bar{g}-3\bar{\zeta})+\chi_g&=0\\ p+\xi-(h/k)_1(\bar{p}+\bar{\xi})+\chi_p&=0\\ q+\xi-(h/k)_1(\bar{q}+\bar{\xi})+\chi_q&=0 \end{aligned}\right\} \quad\ldots\ldots(10)$$

where the correcting $\phi_4(z)$ function is represented by

$$\left.\begin{aligned} \zeta &\equiv \int_l \frac{h}{\sqrt{\Phi}} 2\phi_4 h^3\,dz, & \xi &\equiv \int_l \frac{h}{\sqrt{\Phi}} \phi_4 6h^2 M_1\,dz, \\ \bar{\zeta} &\equiv \int_l \frac{k}{\sqrt{\Phi}} 2\phi_4 h^3\,dz, & \bar{\xi} &\equiv \int_l \frac{k}{\sqrt{\Phi}} \phi_4 6h^2 M_1\,dz, \end{aligned}\right\} \quad \ldots\ldots(11)$$

and similarly, using up to M_2, eqns. (7) and (8) become

$$\left.\begin{aligned} f-(h/k)_l(\bar{f}+\sqrt{\Phi_0})+\chi_f &= 0 \\ d-(h/k)_l\bar{d} \qquad\qquad +\chi_d &= 0 \end{aligned}\right\} \quad \ldots\ldots(12)$$

The suffix l indicates that the integral has only to be taken over the region of the first three electrodes, and the symbols I, e, g, p, q, d and f represent the results of the preliminary integrations, again taken only over that region.

The symbols χ represent contributions from the fourth electrode, and contain the integrals

$$C \equiv \int_C \phi_2\,dz, \qquad S \equiv \int_C (z-z_C)^2 \phi_2\,dz, \quad \ldots\ldots(13)$$

taken only over the region of the fourth electrode (as indicated by the suffix C). This part of the integration has therefore been reduced to the evaluation of C, the 'strength' of the fourth electrode, and of S, which is a less important term. To obtain relatively simple expressions, the region over which the fourth electrode acts has been assumed to be 'thin', so that the alteration of tracks caused by this electrode does not become appreciable within this region. It has also been assumed to be longitudinally symmetric about its centre, so that terms such as $\int_C (z-z_C)\phi_2\,dz$ are zero. The assumption of thinness means that the condition (9) for no distortion may now be replaced by $(M_1+M_3+M_5+\ldots)_C=0$, or, to the suggested degree of approximation,

$$(M_1)_C = 0. \quad \ldots\ldots(14)$$

The interpretation of this is that the fourth-electrode position is as predicted in § 3.

Evaluation of the contributions l may also be reduced to a few initial integrations, by using standard peak shapes $\Omega(z)$ (fig. 2) so that

$$\phi_2(z) = m\Omega_m(z) + n\Omega_n(z). \quad \ldots\ldots(15)$$

m, n are the multiplying constants or 'ellipticities' and suffices m, n indicate positions at the first and third electrodes respectively. Then d becomes

$$d = d_m m + d_n n, \quad \ldots\ldots(16)$$

where

$$d_m \equiv \int_m \frac{h^2\Omega_m}{\sqrt{\Phi}}\,dz \qquad d_n \equiv \int_n \frac{h^2\Omega_n}{\sqrt{\Phi}}\,dz$$

and similarly for p and q, while f, e and g take forms such as

$$f = f_m m^2 + f_{mn} mn + f_n n^2. \quad \ldots\ldots(17)$$

Because no integrand contributions come from the region beyond the lens, these symbols d_m etc. are simply numbers, and may all be evaluated once the standard peak shape has been chosen. Correction conditions are then to be satisfied by algebraic manipulation of the five parameters representing the two 'ellipticities', the positions of fourth-electrode and image, and the fourth-electrode 'strength' (m, n, z_C, z_i, C), and two third-order correction parameters involved in (11).

The seven conditions (10), (12), (14) determine values of these seven parameters which will satisfy the correction conditions. They are not easily solved explicitly, but it may be shown (Burfoot 1952 b) that the S terms are negligible and that for practical high magnification lenses the $(h/k)_1$ terms in (10) and (A 6) may be neglected. The solutions then are

$$n=(p_m-q_m)\times\left[\frac{-4I}{(3e_m+g_m)(p_n-q_n)^2-(3e_{mn}+g_{mn})(p_m-q_m)(p_n-q_n)+(3e_n+g_n)(p_m-q_m)^2}\right]^{1/2} \qquad \ldots\ldots(18)$$

$$m=-n(p_n-q_n)/(p_m-q_m) \qquad \ldots\ldots(19)$$

$$\zeta=-I-e-\tfrac{3}{4}Ch_C{}^2h_1'M'/\sqrt{\Phi_C} \qquad \ldots\ldots(20)$$

$$\xi=-p-\tfrac{1}{2}Ch_C{}^2h_1'^2/\sqrt{\Phi_C} \qquad \ldots\ldots(21)$$

$$z_C=z_b-k_b/(k_1'-h_1'\bar{d}/d) \qquad \ldots\ldots(22)$$

$$C=-d\sqrt{\Phi_C}/h_C{}^2 \qquad \ldots\ldots(23)$$

$$z_1=z_b-k_b/\{k_1'-h_1'(\bar{f}+\sqrt{\Phi_O})/f\}. \qquad \ldots\ldots(24)$$

The solution procedure is summarized at the end of Appendix I, for any given unipotential rotationally symmetric distribution and for any chosen standard peak shapes and object position. (The suffix b represents the position of zero of h.)

§ 5. EXAMPLE AND TOLERANCES

Calculations have been carried out on the rotationally symmetric basis of a typical objective lens, the data for which have been given elsewhere (Burfoot 1952 a)*. This is a unipotential lens with focal length 3 mm and electrode spacing about 2·5 mm, the object position being such as to provide a magnification of 86. The image is about 40 cm from the lens and the spherical aberration constant Cf has the typical value of 2 cm. It may be taken that the lens is used at 100 kv. The author has developed a method for the initial ray tracing which is very simple, and which may be carried to any degree of accuracy within the limits of the data without increasing the complexity. It is found that the additional electrode must be 8 mm from the lens centre. A ray from the object centre is given in fig. 3. The image is found to be 3 cm from the lens centre, and the final magnification is only 6·4. Since the minimum number of controls has been used, there is no freedom of design available to improve this. Presumably different examples could be found with higher magnifications. Alternatively, for some purposes it may be possible to omit the stipulation of no distortion, and this will allow one degree of manipulation with no increase in the number of controls. The theory has been extended to cover such cases—resolution is not then impaired, but central circles are imaged as ellipses; applying the theory to this example, it is found that if the fourth electrode is situated 5 mm from the lens centre, the magnifications are increased to 13 and 8, in the x and y directions.

It is calculated that if the quotient (aberration/magnification) is to be reduced to less than 10% of its value in the original rotationally symmetric basic lens, the first-order parameters m, n, z_C, C must not vary from their calculated values by more than one part in 2×10^4, nor the two third-order parameters by more than one in 10^3. It is verified that the approximations made to enable easy solution of the seven equations cannot introduce errors of this magnitude. It is not thought worth while to repeat the whole solution relativistically, though it must be recognized that non-relativistic ray tracing from a given object position is not a good approximation to physical rays from that position at these voltages in practical high magnification lenses. The focal length is about 2% greater than its non-relativistic estimate, so that the object position which physically corresponds to the above image position is about fifty microns further from the lens than the position which has been taken non-relativistically. The former is the position which would automatically be adopted by a microscope operator in focusing upon the screen, and it is found that a physical ray from this object centre differs from the non-relativistic trace from the non-relativistic position by not more than 5% of the off-axial distance, even in those regions largely responsible for the aberrations. It is therefore believed that the non-rotational symmetry parameters are not very critical with respect to the object position and that the principles of the correction cannot be greatly affected by the non-relativistic nature of the calculations, though the numerical values of the parameters may be slightly in error. For lens design, relativistic working would be essential, but in this paper we are concerned only with their tolerances, which will clearly not be very different. The validity of working only with the first two modifications *M* and of treating the fourth aperture region as 'thin' has been justified by adapting the ray tracing method to carry out direct ray tracing through the lens finally determined. Ray tracing is necessarily equivalent to using all the *M* terms, and it is found that the traces never differ from those of fig. 3 by as much as 1%. For completeness, it must be pointed out that in achieving a degree of spherical correction, it is no longer necessarily true that spherical aberration is predominant among the third-order aberrations. However, no other aberration can affect the centre of the image, and in the interests of a high resolution there, large fields of view might often be willingly sacrificed, as is the case with astronomical telescopes.

---

* The unit of length may there be taken as 0·2 mm.

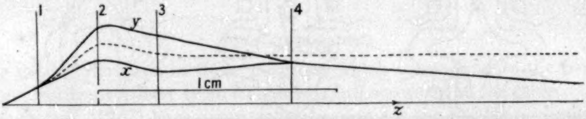


Fig. 3. x and y projections of a ray from the object centre, with arbitrary ordinates, not in general the same for both projections. The broken line is the corresponding ray in the basic rotationally symmetric lens.

Finally, it is proposed to determine a corresponding electrode structure in order to be able to examine the manufacturing tolerances which arise. It is well known that very different potential distributions away from the axis can produce essentially the same distributions near the axis, and the same aberrations with small apertures. A desirable axial distribution has been determined above, but no general method exists for enumerating the large group of distributions to which this belongs. Therefore one of the group has been taken arbitrarily. The distribution chosen is one which it is found can be calculated without using either infinite series such as (4) or summation over a range of values of a separation parameter *K*. It is discussed in Appendix II. The resulting distribution at the first aperture may take the form sketched in fig. 4(*a*), or for stronger departures from rotational symmetry, that of 4(*b*), in which the x, y cross section is an exact plot corresponding to the first electrode in the above example. In the case 4(*b*) the equipotential surfaces with potentials above 100 kv are closed surfaces with two sharp points on each where they are cut by the circle $r=R$, $z=z_m$ or $\nu=\mu=0$. Those below 100 kv form the surfaces of apertures in thick plates, the lower potentials dividing into two apertures. At still lower potentials these two apertures become 'sealed over' on both faces and separate into two closed surfaces and two near-planes. Finally the lowest potentials consist simply of two planes $z=$ constant at increasing distances from the 'aperture plane'.

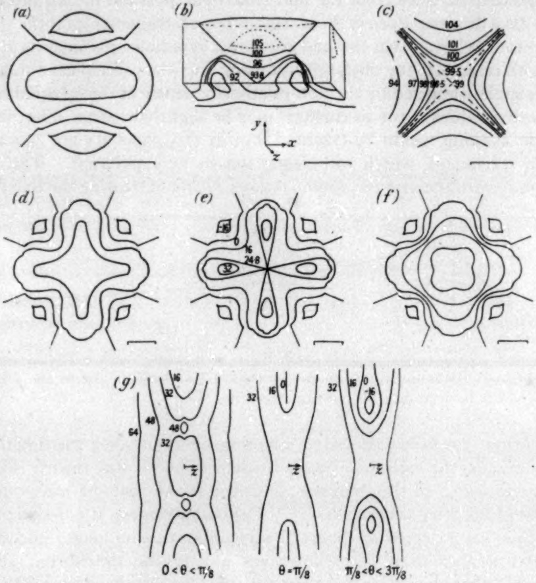


Fig. 4. A method of realizing the axial requirements. Forms of the equipotential surfaces. For details see text. The figures indicate kilovolts; cathode potential = 0. The scale markers are ½ mm.

The third aperture is similar, but rotated about the axis through 90°. The distribution at the fourth aperture possesses only the ϕ_2 component, and is shown exactly in cross section in fig. 4(*c*) for the example above. The cross section of fig. 4(*b*) combines the characteristics of 4(*c*) and the concentric circles which represent the cross section of the rotationally symmetric components. The two parameters of ϕ_4 may be applied similarly at the second aperture. Since the two independent peaks are applied superimposed, in this case the actual shape of the composite peak will vary somewhat as the parameters vary, while the cross section may take any of the forms sketched in figs. 4(*d*), (*e*), (*f*), this last being a case where the two terms are peaks of opposing sign. Typical profiles for 4(*e*) are also sketched in 4(*g*).

Chosen equipotential surfaces from these distributions must be fashioned in conducting material to provide the physical electrodes, and the accuracy with which these surfaces would have to be produced must be such as to cause no greater inaccuracy in the required parameter values than one in 2×10^4 or one in 10^3 as appropriate, to improve the resolution limit by a factor of ten. The best choice from among these surfaces seems to be to select those whose potential is that of the original rotationally symmetric electrodes, 100 kv and 0 (cathode potential). These consist of tongue shaped surfaces (the *yz* profile of which has also been indicated in fig. 4(*b*)). The group of surfaces to one side of this choice would consist of closed surfaces involving sharp points, while the other group approaches nearer to the axis, and must present more exacting tolerances. In the case of the fourth aperture it would be necessary to fashion, say, the 100 kv surface and its 'twin' 98 kv. Such complicated shapes are not, of course, suggested as a practical lens design, but they may give indications applicable to more practical methods of realizing the required axial distribution, if these can be discovered. In particular, the required tolerances may be estimated. A general theory would be complex, but the positions of the tips of the chosen tongues, where the electrodes approach nearest to the axis, may easily be investigated.

§ 6. CONCLUSIONS

The radial tolerances in these positions which correspond to the parameter variations mentioned, have been calculated, and are given in the table. Bertein and Regenstreif (1947) have given an expression for the resolution limit caused by a slight unintentional ellipticity of the second aperture of a unipotential lens. It may be interesting to compare the order of magnitude of this ellipticity with that of the first-order effects in the table, 40 Å and 100 Å. Applying their expression, a limit of 10% of 10 Å would occur, in a lens with semi-aperture angle 5×10^{-3} radian, if the semi-axes of the ellipse differed by 200 Å.

Tolerances of Tonguetips

First electrode	Δr	= 100 Å
Second electrode	Δr	= $\frac{1}{5}$ micron
Third electrode	Δr	= 40 Å
Fourth electrode	Δr	= 100 Å
	Δz_C	= 2 microns

The tolerances in the table are beyond the precision which could be attained in practice, even for much simpler shapes. This merely emphasizes that more practical realizations must be found. For example, it is possible to develop expressions (Burfoot 1952 b, Cotte 1949) for the ϕ_2 and ϕ_4 peaks caused by apertures which are 'strong' ellipses. The difficulty in their application is that these peaks, being each dependent on the semi-axes *a* and *b*, are not independent in shape, nor are they independent of $\Phi(z)$. The author has shown that even maximum ellipticity (parallel-sided slit) will only provide astigmatisms a factor of ten smaller than those required, if $\Phi(z)$ is to remain fixed.

It is to be observed that the second electrode, which provides the specifically third-order correction, is less exacting than the first-order electrodes by one to two factors of ten. A different example might bring the tolerance of the second electrode within reach, but it seems improbable that the first-order electrodes could be dealt with similarly. It may be concluded that it is not impossible, in principle, to design lenses with 'built-in' spherical correction, and that this can be done with a small number of electrodes while the calculations still remain feasible, that a strong astigmatism must be produced as an auxiliary effect, and that it is the subsequent removal of this which constitutes the primary challenge to engineering ingenuity.

Screw-threads can be manufactured more accurately than complicated electrode shapes; and neighbouring equipotential surfaces are similar to one another in shape. It seems possible, therefore, that a system could be designed, similar to that discussed here, but in which final correction is achieved empirically by the provision of radial position-adjustments and fine potential adjustments (say, 1 part in 10^4 or better). The necessary strength of astigmatism might be produced by fixed electrodes which approach close to the axis, with a 'fine control' provided by additional adjustable electrodes at a greater distance.

The foregoing theory is limited to the field of geometrical optics, since the present degree of lens perfection does not seem yet to warrant the introduction of more exact 'wave' theory or consideration of the effects of mutual repulsions between electrons. Electron optics is likely to remain primarily an empirical and practical science, but it has been shown here that, nevertheless, some theoretical guidance may be given, at least to indicate where the difficulties are likely to arise.

ACKNOWLEDGMENTS

This work was carried out at the Cavendish Laboratory under the supervision of Dr. V. E. Cosslett and with the aid of a maintenance grant from the Ministry of Education.

APPENDIX I*

The solutions of $L=0$ may be written as

$$\left.\begin{aligned} x_\alpha &= \frac{x_A}{(h_x)_A} h_x(z) + x_0 k_x(z) \\ y_\alpha &= \frac{y_A}{(h_y)_A} h_y(z) + y_0 k_y(z) \end{aligned}\right\} \quad \ldots\ldots (\mathrm{A}\,1)$$

where h_x, h_y are solutions which are zero and with slopes unity at the object plane z_0, while the values of k_x, k_y are unity at z_0; $(h_x)_A$, $(h_y)_A$, are the values of h_x, h_y at the longitudinal position z_A given by $k_x=0$, $k_y=0$ respectively (fig. 1). Since k is only partially specified, the positions z_A are arbitrary. x_0, y_0, x_A, y_A characterize the particular solution represented by (A 1), being the values of this solution at z_0 and z_A. The first-order image plane for the x coordinates of the electron

---

[*] The more lengthy expressions are collected for convenience in this Appendix, which should be read in conjunction with the text.

path occurs at that value of z, other than z_O, where $h_x(z)$ is zero, the magnification in the x direction being given by the value of $k_x(z)$ or $1/h_x'(z)$ (using Helmholtz' theorem, see § 3) at that plane, and similarly for the y coordinates. The conditions for perfect imaging of each point of the object are therefore that the zeros of h_x, h_y should occur at the same position z, and that the integrals representing the third order aberration should vanish at the same position. If in addition there is to be no first-order distortion, it must also be ensured that $h_x'(z)=h_y'(z)$ at that position.

In terms of (A 1), the exact solution of $L=F$ at any position z is

$$x(y,z)=x_\alpha+\int_O^z [x,y]\,dz \qquad y(x,z)=y_\alpha+\int_O^z [y,x]\,dz, \quad \ldots\ldots(\text{A } 2)$$

where

$$\int_O^z [x,y]\,dz \equiv \frac{h_x}{\{\Phi_O(1+\kappa\Phi_O)\}^{1/2}}\int_O^z \{\Phi(1+\kappa\Phi)\}^{1/2}Fk_x\,dz - \frac{k_x}{\{\Phi_O(1+\kappa\Phi_O)\}^{1/2}}\int_O^z \{\Phi(1+\kappa\Phi)\}^{1/2}Fh_x\,dz$$

and similarly for $\int_O^z [y,x]\,dz$ in the y equation. Successive approximations to this may be written as

$$x_\beta=x_\alpha+\int_O^z [x_\alpha,y_\alpha]\,dz \qquad y_\beta=y_\alpha+\int_O^z [y_\alpha,x_\alpha]\,dz$$

$$x_\gamma=x_\alpha+\int_O^z [x_\beta,y_\beta]\,dz \qquad y_\gamma=y_\alpha+\int_O^z [y_\beta,x_\beta]\,dz$$

with an obvious extension of the notation.

The non-astigmatic parts of x_α, y_α which are used as the first approximate solution in this paper may be denoted by

$$\bar{x}=\frac{x_A}{h_A}h(z)+x_Ok(z) \qquad \bar{y}=\frac{y_A}{h_A}h(z)+y_Ok(z). \quad \ldots\ldots(\text{A } 3)$$

(The arbitrary position z_A must be chosen the same for both x and y; see fig. 1) and in terms of $h(z)$ and $k(z)$, the first-order astigmatic solutions (A 1) take the form of (6), by applying (A 2) to the transferred ϕ_2 term. In (6)

$$M_1(z)\equiv h\int_O^z hkB\,dz-k\int_O^z h^2B\,dz \qquad M_1{}^O(z)\equiv h\int_O^z k^2B\,dz-k\int_O^z khB\,dz$$

$$\cdots\cdots\cdots\cdot \qquad\qquad \cdots\cdots\cdots\cdot$$

$$M_{p+1}(z)\equiv h\int_O^z M_pkB\,dz-k\int_O^z M_phB\,dz$$

$$M_{p+1}{}^O(z)\equiv h\int_O^z M_p{}^OkB\,dz-k\int_O^z M_p{}^OhB\,dz$$

with

$$B(z)\equiv\left[\frac{\Phi(1+\kappa\Phi)}{\Phi_O(1+\kappa\Phi_O)}\right]^{1/2}\frac{\phi_2}{\Phi}\frac{1+2\kappa\Phi}{1+\kappa\Phi}.$$

The non-relativistic expressions are

$$M_1(z)\equiv\frac{h}{\sqrt{\Phi_O}}\int_O^z\frac{hk\phi_2}{\sqrt{\Phi}}\,dz-\frac{k}{\sqrt{\Phi_O}}\int_O^z\frac{h^2\phi_2}{\sqrt{\Phi}}\,dz \text{ etc.} \quad \ldots\ldots(\text{A } 4)$$

Retaining the most significant terms as suggested in the text, the spherical aberration terms are

$$\delta x=\frac{x_A{}^3}{\sqrt{\Phi_0}}\left[h\int_0^i\frac{k}{\sqrt{\Phi}}(W_e+W_p)dz-k\int_0^i\frac{h}{\sqrt{\Phi}}(W_e+W_p)dz\right]$$
$$+\frac{x_Ay_A{}^2}{\sqrt{\Phi_0}}\left[h\int_0^i\frac{k}{\sqrt{\Phi}}(W_g+W_q)dz-k\int_0^i\frac{h}{\sqrt{\Phi}}(W_g+W_q)dz\right]$$
$$\delta y=\frac{y_A{}^3}{\sqrt{\Phi_0}}\left[h\int_0^i\frac{k}{\sqrt{\Phi}}(W_e-W_p)dz-k\int_0^i\frac{h}{\sqrt{\Phi}}(W_e-W_p)dz\right]$$
$$+\frac{y_Ax_A{}^2}{\sqrt{\Phi_0}}\left[h\int_0^i\frac{k}{\sqrt{\Phi}}(W_g-W_q)dz-k\int_0^i\frac{h}{\sqrt{\Phi}}(W_g-W_q)dz\right]$$

where

$$W_e\equiv W_{e0}+W_{e2}+2\phi_4h^3$$
$$W_g\equiv W_{g0}+W_{g2}-6\phi_4h^3$$
$$W_p\equiv \qquad W_{p2}+\phi_4(6h^2M_1)$$
$$W_q\equiv \qquad W_{q2}+\phi_4(6h^2M_1)$$

with

$$W_{e0}=W_{g0}\equiv\frac{\Phi''}{4}h^2h''+\frac{\Phi^{(4)}}{32}h^3-\frac{\Phi''}{4}hh'^2+\frac{\Phi'''}{8}h^2h'-\frac{\Phi'}{2}h'^3$$
$$W_{e2}\equiv\phi_2(-2hh''M_1-h^2M_1''+2hh'M_1'+h'^2M_1)+\phi_2'(-hh'M_1-\tfrac{1}{2}h^2M_1')$$
$$+\phi_2''(-\tfrac{1}{2}h^2M_1)$$
$$W_{g2}\equiv\phi_2(-2hh''M_1+h^2M_1''-4hh'M_1'+h'^2M_1)+\phi_2'(-hh'M_1+\tfrac{1}{2}h^2M_1')$$
$$W_{p2}\equiv-\phi_2h^2h''-\tfrac{1}{2}\phi_2'h^2h'+\phi_2hh'^2-\tfrac{1}{6}\phi_2''h^3$$
$$+\Phi'(-\tfrac{3}{2}h'^2M_1')+\Phi''(\tfrac{1}{2}hh''M_1+\tfrac{1}{4}h^2M_1''-\tfrac{1}{4}h'^2M_1-\tfrac{1}{2}hh'M_1')$$
$$+\Phi'''(\tfrac{1}{4}hh'M_1+\tfrac{1}{8}h^2M_1')+\Phi^{(4)}(\tfrac{3}{32}h^2M_1)$$
$$W_{q2}\equiv\phi_2h^2h''+\tfrac{1}{2}\phi_2'h^2h'+3\phi_2hh'^2+\Phi(2h'h''M_1'-2h'^2M_1'')+\Phi'(+\tfrac{1}{2}h'^2M_1')$$
$$+\Phi''(-\tfrac{1}{2}hh''M_1+\tfrac{1}{4}h^2M_1''-\tfrac{3}{4}h'^2M_1+hh'M_1')$$
$$+\Phi'''(-\tfrac{1}{4}hh'M_1+\tfrac{1}{8}h^2M_1')+\Phi^{(4)}(-\tfrac{1}{32}h^2M_1) \qquad \ldots\ldots(\text{A }5)$$

from which it may be seen that the spherical aberration correction conditions are

$$h_i\int_0^i\frac{k}{\sqrt{\Phi}}W_jdz-k_i\int_0^i\frac{h}{\sqrt{\Phi}}W_jdz=0 \qquad (j=e,g,p,q).$$

Being obtained from those terms of (6) which only involve x_A and y_A, M^0 terms do not occur so there are no 'crossed terms' such as $M_1M_1{}^0$. These equations and eqns. (7)–(8) may be written as (10) and (12) where

$$\chi_e\equiv[\tfrac{3}{4}h_i'M'(h_C{}^2C+h_i'^2S)-(h/k)_i\{h_CM'C(h_i'k_C-\tfrac{1}{4}h_Ck_i')+\tfrac{3}{4}h_i'^2k_i'M'S\}]/\sqrt{\Phi_C}$$
$$\chi_g\equiv[-\tfrac{3}{4}h_i'M'(3h_C{}^2C+h_i'^2S)-(h/k)_i\{-h_CM'C(2h_i'k_C+\tfrac{1}{4}h_Ck_i')$$
$$-\tfrac{3}{4}h_i'^2k_i'M'S\}]/\sqrt{\Phi_C}$$
$$\chi_p\equiv[\tfrac{1}{2}h_i'^2(h_C{}^2C+h_i'^2S)-(h/k)_i\{h_Ch_i'(h_i'k_C-\tfrac{1}{2}h_Ck_i')C+\tfrac{1}{2}h_i'^3k_i'S\}]/\sqrt{\Phi_C}$$
$$\chi_q\equiv[-\tfrac{1}{2}h_i'^2(h_C{}^2C+h_i'^2S)-(h/k)_i\{-\tfrac{1}{2}h_i'k_i'(h_C{}^2C+h_i'^2S)\}]/\sqrt{\Phi_C}$$
$$\chi_f\equiv[\tfrac{1}{2}h_i'M'S-(h/k)_i\tfrac{1}{2}k_i'M'S]/\sqrt{\Phi_C}$$
$$\chi_d\equiv[h_C{}^2C+h_i'^2S-(h/k)_i(h_Ck_CC+h_i'k_i'S)]/\sqrt{\Phi_C} \qquad \ldots\ldots(\text{A }6)$$

and where

$$\left.\begin{aligned} I&\equiv\int_l \frac{h}{\sqrt{\Phi}}W_{e0}\,dz & j&\equiv\int_l \frac{h}{\sqrt{\Phi}}W_{j2}\,dz \\ \bar{I}&\equiv\int_l \frac{k}{\sqrt{\Phi}}W_{e0}\,dz & \bar{j}&\equiv\int_l \frac{k}{\sqrt{\Phi}}W_{j2}\,dz \\ f&\equiv\int_l \frac{M_1h\phi_2}{\sqrt{\Phi}}\,dz & d&\equiv\int_l \frac{h^2\phi_2}{\sqrt{\Phi}}\,dz \\ \bar{f}&\equiv\int_l \frac{M_1k\phi_2}{\sqrt{\Phi}}\,dz & \bar{d}&\equiv\int_l \frac{hk\phi_2}{\sqrt{\Phi}}\,dz \end{aligned}\right\} \quad\ldots\ldots\text{(A 7)}$$

M' is written for the value of $M_1'(z)$ in the field-free region preceding the fourth electrode. I and $\bar{I}$ represent the aberration of the fixed basic rotationally symmetric lens, which by Scherzer's theorem is non-zero, while h_i', k_i' are also constants characteristic of that lens. The 'control' z_C has been taken outside the integrals (C and S), being implicit in the symbols

$$h_C=h_i'(z_C-z_b), \qquad k_C=k_b+k_i'(z_C-z_b).$$

Similarly z_i occurs as $(h/k)_i$.

Solution procedure for given $\Phi(z)$, $\Omega(z)$ and z_O: Determine $h(z)$, $k(z)$ by numerical ray tracing methods, read off the values of the constants h_i', k_i', k_b, z_b, and determine the m, n components of $M_1(z)$ (using (A 4) with ϕ_2 replaced by Ω_m and Ω_n), and hence find the m, n components of W_{j2} from (A 5). Carry out the appropriate preliminary integrations to evaluate $I, f_m, d_m, e_m, g_m, p_m, q_m$, etc., and then apply successively eqns. (18), (19), (16), (17), etc., and (22). Read off Φ_C from the $\Phi(z)$ distribution and apply (23) using $h_C=-k_b/(k_i'/h_i'-\bar{d}/d)$. Finally apply (20)–(21), using $M'=(h_i'\bar{d}-k_i'd)/\sqrt{\Phi_O}$. Equation (24) gives the image position at which the corrected image will be found, and it may be shown that the magnification of this image is $\Phi_O/[(\bar{f}h_i'-fk_i')\sqrt{\Phi_i}]$. It is then a simple matter to find what peaks of $\phi_4(z)$ applied at the second aperture in (11) give the calculated values for ζ and ξ.

APPENDIX II

The series (4) was obtained by substituting an infinite series in r into Laplace's equation. By substituting instead a series in θ, the ϕ_2 term can be selected in the form $\cos 2\theta\Sigma_K A_K f(K,u)g(K,v)$ where (u,v,θ) are the coordinates used. It is found however that by suitable choice of u, v, one value only of K is sufficient to provide function peaks of a suitable shape, avoiding the summation.

Consider the oblate spheroidal coordinates (μ, ν, θ) centred on one of the electrode positions z_m and related to (r, z, θ) by $r^2=R^2(1+\mu^2)(1-\nu^2)$, $z-z_m=R\mu\nu$, with the scale factor R equal to the radius of the original circular aperture in that electrode. It can be shown that Laplace's equation in these coordinates is

$$\frac{\partial}{\partial\mu}\left\{(\mu^2+1)\frac{\partial\phi}{\partial\mu}\right\}+\frac{\partial}{\partial\nu}\left\{(1-\nu^2)\frac{\partial\phi}{\partial\nu}\right\}+\frac{\partial}{\partial\theta}\left\{\frac{\nu^2+\mu^2}{(1+\mu^2)(1-\nu^2)}\frac{\partial\phi}{\partial\theta}\right\}=0,$$

and if we substitute $\phi=F_0+F_2\cos 2\theta+F_4\cos 4\theta+\ldots+F_p\cos p\theta+\ldots$ in which F are functions of μ, ν we find

$$F_p=\int\{A_KP_K{}^p(\nu)+B_KQ_K{}^p(\nu)\}\{C_KP_K{}^p(i\mu)+D_KQ_K{}^p(i\mu)\}\,dK,$$

where $i\equiv\sqrt{-1}$ and where $P_K{}^p$, $Q_K{}^p$ are the associated Legendre functions of the first and second kinds respectively which have singularities at ±1. K in general takes all complex values, but finite solutions at $\nu = \pm 1$ can occur only if K^2 has the form $K^2 = l(l+1)$ with $l = p, p+1, p+2, \ldots$, and if the second kind solution is excluded. So all B_K must be zero and the integrals simplify to sums. Also all C_K must be zero because $P_K{}^p(i\mu)$ tends to infinity as μ becomes large. Thus putting $D_K = 1$ without loss of generality, we find

$$\phi = \sum_{l=0}^{\infty} A_{0,l} P_l(\nu) Q_l(i\mu) + \cos 2\theta \sum_{l=2}^{\infty} A_{2,l} P_l^2(\nu) Q_l^2(i\mu)$$
$$+ \cos 4\theta \sum_{l=4}^{\infty} A_{4,l} P_l^4(\nu) Q_l^4(i\mu) + \ldots .$$

Comparison with (A 3) will show that

$$\phi_p(z) = \sum_{l=p}^{\infty} \frac{A_{p,l}}{R^p} T_{p,l} \frac{Q_l^p}{\{1+(z-z_m)^2/R^2\}^{p/2}}, \qquad \ldots\ldots(\text{A } 8)$$

where the numbers T are $[P_l^p(\nu)/(1-\nu^2)^{p/2}]_{\nu=1}$ and where the argument of Q is now $i(z-z_m)/R$. It may be noted here that the term $p=0$, $l=1$ can be shown to lead to the expression of Regenstreif (1949) from which the basic lens is taken. His expression is a special case of the following, which is in 'mixed' coordinates μ and z

$$\phi = \alpha + \beta[(\tan^{-1}\mu_m + 1/\mu_m)|z+\gamma| - 2(\tan^{-1}\mu_a + 1/\mu_a)|z|$$
$$+ (\tan^{-1}\mu_n + 1/\mu_n)|z-\gamma|] - \beta(R_m - 2R_a + R_n).$$

Here α, β are Regenstreif's constants and suffices m, a, n indicate first, second, or third, electrode, situated at $z = -\gamma$, $z = 0$, $z = \gamma$ respectively. The functions of (A 8) (multiplied by i^{l+2p+1}, since they are not all real) increase from zero at large positive values of z to large positive values at negative z, but they cross the ordinate axis at a finite angle, and the sign ambiguity of the μ coordinate (e.g. Margenau and Murphy 1943) allows us to 'reflect' these functions in the region of negative z. The resulting functions are individually suitable for the standard peak shapes. The summations of (A 8) can therefore be avoided by using, for example in the case of ϕ_2, only one of the terms from the function $p=2$ (say $l=2$). Thus taking $A_{2,2} = 1$, the standard peak shape in (15) may be taken as

$$\Omega(z) = (-3i/R) Q_2{}^2 \left\{\frac{i(z-z_m)}{R}\right\} \Big/ \left\{1 + \frac{(z-z_m)^2}{R^2}\right\}.$$

For the two ϕ_4 peaks, two terms must be used from the function $p=4$ in (A 8).

References

Bertein, F., and Regenstreif, E., 1947, *C. R. Acad. Sci., Paris*, **224**, 737.

Burfoot, J. C., 1952 a, *Brit. J. Appl. Phys.*, **3**, 22; 1952 b, *Thesis*, University of Cambridge.

Cotte, M., 1949, *C. R. Acad. Sci., Paris*, **228**, 377.

Hillier, J., 1946, *J. Appl. Phys.*, **17**, 307.

Margenau, H., and Murphy, G. M., 1943, *The Mathematics of Physics and Chemistry* (New York : Van Nostrand), p. 175.

Rebsch, R., 1938, *Ann. Phys., Lpz.*, **31**, 551.

Regenstreif, E., 1949, *C. R. Acad. Sci., Paris*, **229**, 1311.

Scherzer, O., 1936, *Z. Phys.*, **101**, 593; 1947, *Optik*, **2**, 114; 1949, *J. Appl. Phys.*, **20**, 20.

Seeliger, R., 1949, *Optik*, **5**, 490.

Zworykin, V. K., *et al.*, 1945, *Electron Optics and the Electron Microscope* (New York: Wiley), p. 355.

Edge Dislocations in Inhomogeneous Media

By A. K. HEAD
H. H. Wills Physical Laboratory, University of Bristol,
and
Aeronautical Research Laboratories, Melbourne, Australia

Communicated by N. F. Mott ; MS. received 18*th May* 1953

Abstract. A general method is given, by which problems of elastic plane strain in a bimetallic medium can be reduced to standard problems of potential theory. A bimetallic medium is one where the elastic constants are different on either side of a plane. The boundary between the two semi-infinite solids can be either a complete weld, or it can transmit direct stress but not shear stresses. A special case is a semi-infinite solid with a free surface. Specific examples considered are (i) an edge dislocation in a semi-infinite solid with a free surface, (ii) an edge dislocation in a bimetallic solid for both types of boundary. In neither case can the stress field be given a simple interpretation in terms of images. However, the stress tending to move the dislocation is the same as that of a dislocation of suitable strength at the image point.

§ 1. INTRODUCTION

In a previous paper (Head 1953) the interaction of an elastic screw dislocation with an idealized grain boundary was solved using an analogy between screw dislocations and electrostatic line charges. In this paper the corresponding interaction of an edge dislocation is found. A general method is described, by which problems of plain strain in an infinite bimetallic elastic solid can be reduced to two standard problems in potential theory. 'Bimetallic' is a convenient description of a solid in which there are different elastic constants on either side of the plane $x=0$. In the general case the two half spaces are taken to be completely welded together. Other cases which will be considered are a semi-infinite solid with a free boundary at $x=0$, and a bimetallic solid where the boundary can transmit direct stress but not shear stresses. The latter case could be taken as a model of a grain boundary at high temperatures.

Specific examples which will be considered are (i) the interaction of an edge dislocation of arbitrary Burgers vector, with a free surface, (ii) the interaction of an edge dislocation with a welded interface and with a slipping interface.

The notation used is that of Sokolnikoff (1946) and fundamental equations will be given their reference number in this book in the form (S.24.13).

§ 2. GENERAL METHOD

We consider an infinite isotropic elastic medium, with elastic constants E_1, σ_1 for $x>0$ and E_2, σ_2 for $x<0$, under conditions of plane strain in the xy plane. In the following, subscripts 1, 2 will refer to these two regions and the omission of subscripts indicates a relationship true for either region on insertion of the appropriate subscript.

The three types of boundary considered are:

(i) Free boundary. $E_x = 0 = \sigma_x$ for $x < 0$
$\tau_{xx} = 0 = \tau_{xy} = \tau_{xz}$ at $x = 0$.

(ii) Welded boundary. The displacements u, v, w and the stresses τ_{xx}, τ_{xy}, τ_{xz} are continuous across $x = 0$.

(iii) Slipping boundary. $\tau_{xy} = 0 = \tau_{xz}$ at $x = 0$ and u, τ_{xx} continuous across $x = 0$.

We now derive suitable harmonic functions, related to the stresses, for which there are simple boundary conditions at $x = 0$. Firstly, it is known (S.24.13) that, in the absence of body forces, the hydrostatic component of the stress field $\theta = \tau_{xx} + \tau_{yy} + \tau_{zz}$ is a harmonic function. In the case of plane strain considered here, $\theta = (1+\sigma)(\tau_{xx} + \tau_{yy})$.

Next we define a function T, such that

$$T = \tau_{xx} + \frac{1}{2(1+\sigma)} x \frac{\partial \theta}{\partial x}.$$

Then
$$\nabla^2 T = \nabla^2 \tau_{xx} + \frac{1}{1+\sigma} \frac{\partial^2 \theta}{\partial x^2} + \frac{1}{2(1+\sigma)} x \nabla^2 \left(\frac{\partial \theta}{\partial x} \right).$$

Since θ is a harmonic function, $\nabla^2(\partial\theta/\partial x) = 0$ and by the compatibility equations of Beltrami

$$\nabla^2 \tau_{xx} + \frac{1}{1+\sigma} \frac{\partial^2 \theta}{\partial x^2} = 0. \qquad \text{(S.24.17)}$$

Hence $\nabla^2 T = 0$ and since at $x = 0$, $T = \tau_{xx}$ we have in case (i) that $T = 0$ at $x = 0$, and in cases (ii) and (iii) that T is continuous at $x = 0$.

We now construct linear combinations of T and θ, which will therefore be harmonic functions, and for which there are suitable boundary conditions at $x = 0$. We derive the first one from the equation of equilibrium (S.15.3) which can be written

$$\begin{aligned} -\frac{\partial \tau_{xy}}{\partial y} - \frac{\partial \tau_{xz}}{\partial z} &= \frac{\partial \tau_{xx}}{\partial x} \\ &= \frac{\partial}{\partial x}\left\{ T - \frac{\theta}{2(1+\sigma)} \right\} - \frac{x}{2(1+\sigma)} \frac{\partial^2 \theta}{\partial x^2} \qquad \text{......(1)} \end{aligned}$$

by the definition of T and θ. In cases (i) and (iii) the left-hand side of (1) is zero at $x = 0$ and therefore $\partial\{T - \theta/(2+2\sigma)\}/\partial x = 0$ at $x = 0$. In case (ii) the left-hand side of (1) is continuous across $x = 0$ and so therefore is $\partial(T - \theta/(2+2\sigma))/\partial x$.

Similarly the first of Navier's equations of equilibrium (S.24.7) may be written

$$\begin{aligned} -\frac{\partial^2 u}{\partial y^2} - \frac{\partial^2 u}{\partial z^2} &= \frac{\partial^2 u}{\partial x^2} + \frac{1}{E} \frac{\partial \theta}{\partial x} \\ &= \frac{\partial}{\partial x}\left\{ \frac{1+\sigma}{E} T + \frac{1-2\sigma}{2E} \theta \right\} - \frac{x}{2E} \frac{\partial^2 \theta}{\partial x^2}. \qquad \text{......(2)} \end{aligned}$$

In cases (ii) and (iii) the left-hand side of (2) is continuous across $x = 0$ and so therefore is $\partial\{(1+\sigma)T/E + (1-2\sigma)\theta/2E\}/\partial x$.

Finally we consider the quantity $\partial v/dy + \partial w/\partial z$ which can be written $(1-\sigma)\theta/E - (1+\sigma)T/E + (x/2E)\partial\theta/\partial x$. Since, in case (ii), $\partial v/\partial y + \partial w/\partial z$ is continuous across $x = 0$, so also is $(1-\sigma)\theta/E - (1+\sigma)T/E$.

We now collect these results for each type of boundary.

§ 3. FREE BOUNDARY

The boundary conditions at $x=0$ which have been found are $T=0$ and $\partial\{T-\theta/(2+2\sigma)\}/\partial x=0$. If we write $U=T-\theta/(2+2\sigma)$ then the two potential problems to be solved are

(*a*) $\nabla^2 T=0$ with $T=0$ at $x=0$.

(*b*) $\nabla^2 U=0$ with $\partial U/\partial x=0$ at $x=0$.

These are standard boundary value problems of potential theory and in the specific case considered below can be solved by the method of images. Having found T and U by whichever method is suitable, we have successively

$$\theta=2(1+\sigma)(T-U)$$

$$\tau_{xx}=T-\frac{x}{2(1+\sigma)}\frac{\partial\theta}{\partial x},\quad \tau_{yy}=\frac{\theta}{1+\sigma}-\tau_{xx}$$

and τ_{xy} given by the equations of equilibrium

$$\frac{\partial\tau_{xy}}{\partial x}=-\frac{\partial\tau_{yy}}{\partial y};\quad \frac{\partial\tau_{xy}}{\partial y}=-\frac{\partial\tau_{xx}}{\partial x}.$$

§ 4. SLIPPING BOUNDARY

The boundary conditions at $x=0$ are that

$$T \qquad \text{is continuous}$$

$$\frac{\partial}{\partial x}\left\{T-\frac{\theta}{2(1+\sigma)}\right\}=0 \qquad \ldots\ldots(3)$$

$$\frac{\partial}{\partial x}\left\{\frac{1+\sigma}{E}T+\frac{1-2\sigma}{2E}\theta\right\} \qquad \text{is continuous} \qquad \ldots\ldots(4)$$

and on adding $(1+\sigma)(1-2\sigma)/E$ times (3) to (4) we have that

$$\frac{\partial}{\partial x}\left\{\frac{2(1-\sigma^2)}{E}T\right\} \qquad \text{is continuous.}$$

If we write $U=T-\theta/(2+2\sigma)$ and $L=2(1-\sigma^2)/E$ then the two potential problems to be solved are

(i) $\nabla^2 U=0$ with $\partial U/\partial x=0$ at $x=0$.

(ii) $\nabla^2 T=0$ with T and $L\partial T/\partial x$ continuous across $x=0$.

These are two standard boundary value problems of potential theory and the stresses may be found from U and T in the same manner as described for the free boundary.

§ 5. WELDED BOUNDARY

The boundary conditions at $x=0$ are that

$$T \qquad \text{is continuous}$$

$$\frac{1+\sigma}{E}T-\frac{1-\sigma}{E}\theta \qquad \text{is continuous}$$

$$\frac{\partial}{\partial x}\left\{T-\frac{1}{2(1+\sigma)}\theta\right\} \qquad \text{is continuous}$$

$$\frac{\partial}{\partial x}\left\{\frac{1+\sigma}{E}T+\frac{1-2\sigma}{2E}\theta\right\} \qquad \text{is continuous.}$$

Hence
$$\left(\alpha+\frac{1+\sigma}{E}\right)T-\frac{1-\sigma}{E}\theta \qquad \ldots\ldots(5)$$
and
$$\frac{\partial}{\partial x}\left\{\left(\beta+\frac{1+\sigma}{E}\right)T+\left(\frac{1-2\sigma}{2E}-\frac{\beta}{2(1+\sigma)}\right)\theta\right\} \qquad \ldots\ldots(6)$$
are continuous across $x=0$, if α and β are constants which do not change on crossing the boundary.

We now choose α and β so that (6) is proportional to the x derivative of (5) on both sides of the boundary. Then α and β are given by
$$\frac{\alpha+(1+\sigma_1)/E_1}{(1-\sigma_1)/E_1}=\frac{\beta+(1+\sigma_1)/E_1}{\beta/2(1+\sigma_1)-(1-2\sigma_1)/2E_1}=\delta_1 \text{ say} \qquad \ldots\ldots(7)$$
and
$$\frac{\alpha+(1+\sigma_2)/E_2}{(1-\sigma_2)/E_2}=\frac{\beta+(1+\sigma_2)/E_2}{\beta/2(1+\sigma_2)-(1-2\sigma_2)/2E_2}=\delta_2 \text{ say.} \qquad \ldots\ldots(8)$$
On eliminating β between (7) and (8) we find α is given by the quadratic equation
$$\begin{aligned}&\alpha^2\left[\frac{1-2\sigma_1}{(1+\sigma_2)E_1}-\frac{1-2\sigma_2}{(1+\sigma_1)E_2}\right]\\&\quad+\alpha\left[\frac{(1+\sigma_1)(3-4\sigma_1)}{(1+\sigma_2)E_1{}^2}-\frac{(1+\sigma_2)(3-4\sigma_2)}{(1+\sigma_1)E_2{}^2}\right]\\&\quad-\left[\frac{(1+\sigma_1)(3-4\sigma_1)(1-2\sigma_2)}{E_1{}^2E_2}-\frac{(1+\sigma_2)(3-4\sigma_2)(1-2\sigma_1)}{E_1E_2{}^2}\right]=0. \qquad \ldots\ldots(9)\end{aligned}$$
If we denote these two values of α by α' and α'' and the corresponding values of β, δ_1, and δ_2 by the same superscripts and we write
$$V'=\frac{1-\sigma}{E}(\delta'T-\theta) \quad \text{and} \quad K'=\frac{\beta'E}{2(1-\sigma^2)}-\frac{1-2\sigma}{2(1-\sigma)}$$
with corresponding expressions for V'' and K'' then we have (i) $\nabla^2V'=0$ with V' and $K'\partial V'/\partial x$ continuous across $x=0$, and (ii) $\nabla^2V''=0$ with V'' and $K''\partial V''/\partial x$ continuous across $x=0$. From these two potential problems we may find V' and V'' and hence T and θ and, in the same manner as above, the stress distribution.

§ 6. EDGE DISLOCATION AND FREE SURFACE

We here consider the stress field in a semi-infinite homogeneous medium (which extends to the right of the free surface $x=0$) due to an edge dislocation at $(a, 0)$. As there is only one set of elastic constants involved they will be denoted by E, σ. We will consider two cases, one where the Burgers vector of the dislocation is perpendicular to the free surface, and the other where it is parallel to the free surface. If the Burgers vector is at an arbitrary angle to the free surface, then it can be resolved into components perpendicular to and parallel to the free surface (Frank 1951), and the stress field will be the superposition of the stress fields of the components.

§ 7. BURGERS VECTOR PERPENDICULAR TO FREE SURFACE

An edge dislocation, of Burgers vector b parallel to $0x$, at the point $(a, 0)$ in an infinite homogeneous medium, has the following stress field
$$\left.\begin{aligned}\tau_{xx}=-D\frac{y\{3(x-a)^2+y^2\}}{\{(x-a)^2+y^2\}^2} \qquad \tau_{yy}=D\frac{y\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2}\\ \tau_{xy}=D\frac{(x-a)\{(x-a)^2-y^2\}}{\{(x-a)^2-y^2\}^2}\end{aligned}\right\} \quad \ldots\ldots(10)$$

where $D = Eb/4\pi(1-\sigma^2)$

Hence
$$T = D\left\{-\frac{y}{(x-a)^2+y^2} + \frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2}\right\} \qquad \ldots\ldots(11)$$

and
$$U = D\,\frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2} \qquad \ldots\ldots(12)$$

For the same dislocation at the same point in the semi-infinite medium we require that T and U have the same singularities at $(a, 0)$ as (11) and (12) and that $T=0$ and $\partial U/\partial x = 0$ at $x=0$. It will be seen that the solution satisfying these requirements is

$$T = D\left\{-\frac{y}{(x-a)^2+y^2} + \frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2} + \frac{y}{(x+a)^2+y^2} + \frac{2ay(x+a)}{\{(x+a)^2+y^2\}^2}\right\}$$

and

$$U = D\left\{\frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2} - \frac{2ay(x+a)}{\{(x+a)^2+y^2\}^2}\right\}$$

Hence

$$\tau_{xx} = D\left\{-\frac{y\{3(x-a)^2+y^2\}}{\{(x-a)^2+y^2\}^2} + \frac{y\{3(x+a)^2+y^2\}}{\{(x+a)^2+y^2\}^2} + 4axy\,\frac{3\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^3}\right\}$$

$$\tau_{yy} = D\left\{\frac{y\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2} - \frac{y\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^2} + 4ay\,\frac{(2a-x)(x+a)^2+(3x+2a)y^2}{\{(x+a)^2+y^2\}^3}\right\}$$

$$\tau_{xy} = D\left\{\frac{(x-a)\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2} - \frac{(x+a)\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^2} + 2a\,\frac{(a-x)(x+a)^3+6x(x+a)y^2-y^4}{\{(x+a)^2+y^2\}^3}\right\}. \qquad \ldots\ldots(13)$$

These expressions for τ_{xx}, τ_{xy} and τ_{yy} have been written so that the first term corresponds to the stress field of the dislocation and the second term to the field of a dislocation of Burgers vector $-b$ at the image point $(-a, 0)$. These are the terms given by the approximate solution of Koehler (1941). It will be seen that the stress distribution is not just the superposition of the stress fields of the dislocation and an image. The stresses tending to move the dislocation (Eshelby 1951), which are the difference between the actual stresses (13) and the stresses of the dislocation in an infinite medium (10), evaluated at the position of the dislocation, are however equal to those due to a dislocation of opposite sign at the image point $(-a, 0)$.

§ 8. BURGERS VECTOR PARALLEL TO FREE SURFACE

An edge dislocation, of Burgers vector b parallel to $0y$, at the point $(a, 0)$ in an infinite elastic medium has the following stress field

$$\tau_{xx} = D\,\frac{(x-a)\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2} \qquad \tau_{yy} = D\,\frac{(x-a)\{(x-a)^2+3y^2\}}{\{(x-a)^2+y^2\}^2}$$

$$\tau_{xy} = D\,\frac{y\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2}$$

where $D = Eb/4\pi(1-\sigma^2)$.

Hence
$$T = D\left\{-\frac{a}{(x-a)^2+y^2} + \frac{2ay^2}{\{(x-a)^2+y^2\}^2}\right\}$$

and
$$U = D\left\{-\frac{x}{(x-a)^2+y^2} + \frac{2ay^2}{\{(x-a)^2+y^2\}^2}\right\}.$$

For the dislocation in the semi-infinite solid, in the same manner as the previous section, we have

$$T=D\left\{-\frac{a}{(x-a)^2+y^2}+\frac{2ay^2}{\{(x-a)^2+y^2\}^2}+\frac{a}{(x+a)^2+y^2}-\frac{2ay^2}{\{(x+a)^2+y^2\}^2}\right\}$$

$$U=D\left\{-\frac{x}{(x-a)^2+y^2}+\frac{2ay^2}{\{(x-a)^2+y^2\}^2}+\frac{x}{\{(x+a)^2+y^2\}^2}+\frac{2ay^2}{\{(x+a)^2+y^2\}^2}\right\}$$

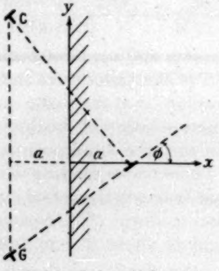
$$\tau_{xx}=D\left\{\frac{(x-a)\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2}-\frac{(x+a)\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^2}\right.$$
$$\left.+2a\,\frac{(3x+a)(x+a)^3-6x(x+a)y^2-y^4}{\{(x+a)^2+y^2\}^3}\right\}$$

$$\tau_{yy}=D\left\{\frac{(x-a)\{(x-a)^2+3y^2\}}{\{(x-a)^2+y^2\}^2}-\frac{(x+a)\{(x+a)^2+3y^2\}}{\{(x+a)^2+y^2\}^2}\right.$$
$$\left.-2a\,\frac{(x-a)(x+a)^3-6x(x+a)y^2+y^4}{\{(x+a)^2+y^2\}^3}\right\}$$

$$\tau_{xy}=D\left\{\frac{y\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2}-\frac{y\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^2}+4axy\,\frac{3(x+a)^2-y^2}{\{(x+a)^2+y^2\}^3}\right\}.$$

§ 9. BURGERS VECTOR AT ARBITRARY ANGLE

If the dislocation has Burgers vector of magnitude b making an angle ϕ with the x axis, then it can be resolved into two dislocations, one with Burgers vector $b\cos\phi$ parallel to the x axis and the other with Burgers vector $b\sin\phi$ parallel to the y axis. Thus the stress field of the dislocation can be obtained by suitable superposition of the two preceding solutions. Of interest is the stress tending to move the dislocation. The 'image' stresses at the position of the dislocation are



$$\tau_{xx}=D\frac{\sin\phi}{2a}\qquad\tau_{yy}=-D\frac{\sin\phi}{2a}\qquad\tau_{xy}=-D\frac{\cos\phi}{2a}.$$

The stress tending to move the dislocation by glide is the shear stress in the direction of the Burgers vector and this is given by

$$\tau_G=(\cos^2\phi-\sin^2\phi)\tau_{xy}-\sin\phi\cos\phi\,(\tau_{xx}-\tau_{yy})$$
$$=-D\,(\cos\phi)/2a.$$

This is the same as the shear stress due to a dislocation of opposite sign in position G of the figure. The stress tending to move the dislocation by climb is the tensile stress in the direction of the Burgers vector and this is given by $\tau_C=-D(\sin\phi)/2a$. This is the same as the tensile stress due to a dislocation of opposite sign in position C of the figure.

§ 10. INTERACTION OF EDGE DISLOCATION AND WELDED BOUNDARY

We consider here the stress field of an edge dislocation at $(a, 0)$ in a bimetallic medium with elastic constants E_1, σ_1 for $x>0$ and E_2, σ_2 for $x<0$. For convenience we consider the case where the direction of the Burgers vector of the dislocation is parallel to $0x$ and of magnitude $b=4\pi^2(1-\sigma_1^2)/E_1$. Thus the constant D of § 7 becomes unity. The stress field for any other value of Burgers vector will be in proportion. The general method of solution has been outlined in § 5, and the same notation is used here.

If this dislocation were in an infinite medium with elastic constants E_1, σ_1 then we would have

$$V'=\frac{1-\sigma_1}{E_1}\{-\delta_1'+2(1+\sigma_1)\}\frac{y}{(x-a)^2+y^2}+\delta_1'\frac{1-\sigma_1}{E_1}\frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2}\quad\ldots\ldots(14)$$

and a similar expression for V'', i.e. with δ_1' replaced by δ_1''.

For the bimetallic solid we require that V should have a singularity of the form (14) and that V and $K\,\partial V/\partial x$ should be continuous at $x=0$. Considering the singularity $y/\{(x-a)^2+y^2\}$ we find that the appropriate function is

$$\frac{y}{(x-a)^2+y^2}+\frac{K_1-K_2}{K_1+K_2}\frac{y}{(x+a)^2+y^2}\quad\text{for}\quad x>0$$

and

$$\frac{2K_1}{K_1+K_2}\frac{y}{(x-a)^2+y^2}\quad\text{for}\quad x<0.$$

Similarly for the singularity $y(x-a)/\{(x-a^2)+y^2\}^2$ the appropriate function is

$$\frac{y(x-a)}{\{(x-a)^2+y^2\}^2}-\frac{K_1-K_2}{K_1+K_2}\frac{y(x+a)}{\{(x+a)^2+y^2\}^2}\quad\text{for}\quad x>0$$

and

$$\frac{2K_1}{K_1+K_2}\frac{y(x-a)}{\{(x-a)^2+y^2\}^2}\quad\text{for}\quad x<0.$$

Thus, for the bimetallic solid, we have

$$V'=\frac{1-\sigma_1}{E_1}\{-\delta_1'+2(1+\sigma_1)\}\left[\frac{y}{\{(x-a)^2+y^2\}}+\frac{K_1'-K_2'}{K_1'+K_2'}\frac{y}{\{(x+a)^2+y^2\}}\right]$$
$$+\delta_1'\frac{1-\sigma_1}{E_1}\left[\frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2}-\frac{K_1'-K_2'}{K_1'+K_2'}\frac{2ay(x+a)}{\{(x+a)^2+y^2\}^2}\right]\quad\text{for}\quad x>0$$
$$=\frac{2K_1'}{K_1'+K_2'}\frac{1-\sigma_1}{E_1}\left[\{-\delta_1'+2(1+\sigma_1)\}\frac{y}{(x-a)^2+y^2}+\delta_1'\frac{2ay(x-a)}{\{(x-a)^2+y^2\}^2}\right]$$
$$\text{for}\quad x<0$$

and a similar expression for V''.
Hence, for the region $x>0$

$$\tau_{xx}=-\frac{y\{3(x-a)^2+y^2\}}{\{(x-a)^2+y^2\}^2}+P\frac{y}{(x+a)^2+y^2}+Q\frac{2a(x+a)y}{\{(x+a)^2+y^2\}^2}$$
$$+\frac{R}{2(1+\sigma_1)}\frac{2x(x+a)y}{\{(x+a)^2+y^2\}^2}-\frac{S}{2(1+\sigma_1)}\left[\frac{2axy}{\{(x+a)^2+y^2\}^2}-\frac{8ax(x+a)^2y}{\{(x+a)^2+y^2\}^3}\right]$$

$$\tau_{yy}=\frac{y\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2}+\left(\frac{R}{1+\sigma_1}-P\right)\frac{y}{(x+a)^2+y^2}+\left(\frac{S}{1+\sigma_1}-Q\right)\frac{2a(x+a)y}{\{(x+a)^2+y^2\}^2}$$
$$-\frac{R}{2(1+\sigma_1)}\frac{2x(x+a)y}{\{(x+a)^2+y^2\}^2}+\frac{S}{2(1+\sigma_1)}\left[\frac{2axy}{\{(x+a)^2+y^2\}^2}-\frac{8ax(x+a)^2y}{\{(x+a)^2+y^2\}^3}\right]$$

$$\tau_{xy} = \frac{(x-a)\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2} - P\frac{x+a}{(x+a)^2+y^2} + Q\left[\frac{a}{(x+a)^2+y^2} - \frac{2a(x+a)^2}{\{(x+a)^2+y^2\}^2}\right]$$
$$+ \frac{R}{2(1+\sigma_1)}\left[\frac{2x+a}{(x+a)^2+y^2} - \frac{2x(x+a)^2}{\{(x+a)^2+y^2\}^2}\right]$$
$$- \frac{S}{2(1+\sigma_1)}\left[\frac{a}{(x+a)^2+y^2} - \frac{2a(x+a)^2+6ax(x+a)}{\{(x+a)^2+y^2\}^2} + \frac{8ax(x+a)^3}{\{(x+a)^2+y^2\}^3}\right]$$

where
$$P = Q + \frac{2(1+\sigma_1)}{\delta_1'-\delta_1''}\left[\frac{K_1'-K_2'}{K_1'+K_2'} - \frac{K_1''-K_2''}{K_1''+K_2''}\right]$$
$$Q = -\frac{1}{\delta_1'-\delta_1''}\left[\delta_1'\frac{K_1'-K_2'}{K_1'+K_2'} - \delta_1''\frac{K_1''-K_2''}{K_1''+K_2''}\right]$$
$$R = S + \frac{2(1+\sigma_1)}{\delta_1'-\delta_1''}\left[\delta_1''\frac{K_1'-K_2'}{K_1'+K_2'} - \delta_1'\frac{K_1''-K_2''}{K_1''+K_2''}\right]$$
$$S = -\frac{\delta_1'\delta_1''}{\delta_1'-\delta_1''}\left[\frac{K_1'-K_2'}{K_1'+K_2'} - \frac{K_1''-K_2''}{K_1''+K_2''}\right].$$

The stress tending to move the dislocation by glide is the value of τ_{xy} at the position of the dislocation, excluding the infinite term corresponding to the dislocation in a homogeneous medium, and is

$$\tau_{xy} = -\frac{1}{2a}\left[P + \frac{Q}{2} - \frac{R}{4(1+\sigma_1)}\right].$$

This is equal to the shear stress which would be produced in a homogeneous medium, by a dislocation at the image point $(-a, 0)$ and of strength $\{R/(4+4\sigma_1) - P - \frac{1}{2}Q\}$ times the strength of the actual dislocation.

For the case $\sigma_1 = \sigma_2$ ($=\sigma$ say) it is found that the constants P, Q, R and S only involve the sum or product of the roots of the quadratic equation for α, eqn. (9), and that the strength of the image dislocation becomes

$$\frac{R}{4(1+\sigma_1)} - P - \tfrac{1}{2}Q = \frac{E_2-E_1}{E_2+E_1}\left\{\frac{E_1}{E_2} + 1 + 2\frac{(1-2\sigma)^2}{3-4\sigma}\right\} \Big/ \left\{\frac{E_1}{E_2} + 1 + 2\frac{(1-2\sigma)^2}{3-4\sigma}\,\frac{2E_1}{E_1+E_2}\right\}.$$

Thus the dislocation is attracted to the interface if $E_2 < E_1$ and repelled if $E_2 > E_1$.

For a screw dislocation under the same conditions (Head 1953) the relative strength of the image is $(E_2-E_1)/(E_2+E_1)$. For $\sigma = 0{\cdot}3$, which is a typical value of Poisson's ratio for metals, the difference between the relative image strengths of screw and edge is less than 15%.

§ 11. SLIPPING BOUNDARY

We consider here the same situation as § 10 except that the interface cannot support shear stresses. Following the method of § 4 we have

$$\tau_{xx} = -\frac{y\{3(x-a)^2+y^2\}}{\{(x-a)^2+y^2\}^2} - M\frac{y\{3(x+a)^2+y^2\}}{\{(x+a)^2+y^2\}^2} + (1-M)\,2axy\,\frac{3(x+a)^2-y^2}{\{(x+a)^2+y^2\}^3}$$
$$\tau_{yy} = \frac{y\{(x-a)^2-y^2\}}{\{(x-a)^2+y^2\}^2} + M\frac{y\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^2}$$
$$+ (1-M)\,2ay\,\frac{(2a-x)(x+a)^2+(3x+2a)y^2}{\{(x+a)^2+y^2\}^3}$$

$$\tau_{xy}=\frac{(x-a)\{(x-a)^2+y^2\}}{\{(x-a)^2+y^2\}^2}+M\frac{(x+a)\{(x+a)^2-y^2\}}{\{(x+a)^2+y^2\}^2}$$
$$+(1-M)a\frac{(a-x)\{(x+a)^3\}+6x(x+a)y^2-y^4}{\{(x+a)^2+y^2\}^3}$$

where $$M=\frac{L_1-L_2}{L_1+L_2}=\left\{\frac{E_2}{1-\sigma_2{}^2}-\frac{E_1}{1-\sigma_1{}^2}\right\}\bigg/\left\{\frac{E_2}{1-\sigma_2{}^2}+\frac{E_1}{1-\sigma_1{}^2}\right\}.$$

The stress tending to move the dislocation by glide is $\tau_{xy}=M/2a$ which is the same as would be produced by a dislocation of strength M at the image point $(-a, 0)$. Thus the dislocation is attracted to, or repelled from, the boundary according as $E_1/(1-\sigma_1{}^2)$ is greater or less than $E_2/(1-\sigma_2{}^2)$.

For this type of boundary the behaviour of an edge dislocation differs from that of a screw dislocation. Since a screw dislocation generates only shear stresses which are not transmitted by the slipping boundary, it will be attracted as if to a free surface. On the other hand, for an edge dislocation, there is little difference between the influence of a slipping boundary and a completely welded boundary.

ACKNOWLEDGMENT

This paper is published by permission of the Department of Supply, Commonwealth of Australia, of which the author is an Officer.

REFERENCES

ESHELBY, J. D., 1951, *Phil. Trans. Roy. Soc.* A, **244**, 87.
FRANK, F. C., 1951, *Phil. Mag.* (7), **42**, 809.
HEAD, A. K., 1953, *Phil. Mag.* (7), **44**, 92.
KOEHLER, J. S., 1941, *Phys. Rev.*, **60**, 397.
SOKOLNIKOFF, I. S., 1946, *Mathematical Theory of Elasticity* (New York : McGraw-Hill).

RESEARCH NOTES

Vertical Focusing in the Microtron

By J. S. Bell
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

Communicated by J. D. Cockcroft ; MS. received 10th June 1953

The energy oscillations in the microtron have been studied in detail (Henderson *et al.* 1953). No such study of the focusing oscillations seems to have been made yet. We consider here the vertical motion only ; this is simplified by the fact that if the static vertical magnetic field is uniform, the only forces to be considered are those arising from the radio-frequency fields in the accelerating gap. Schiff (1946) has remarked that the usual focusing effect of these fields vanishes when the particle travels with the velocity of light. We will show, however, that it does not vanish so rapidly as might be expected, because with the very strong fields in the microtron it is not the usual first order term (which decreases as $1-\beta^2$) which is important, but another term, of second order in field strength, which decreases only as $(1-\beta^2)^{1/2}$.

Let x, y, z be cartesian coordinates with x along the axis of the accelerating gap and z along the static magnetic field. The equation for the z motion is

$$\frac{d}{dt}\left(m\frac{dz}{dt}\right)=\mathbf{e}E_z+\mathbf{e}\beta H_y$$

where m is the relativistic mass and E_z, H_y the components of the radio-frequency field. It is convenient to introduce $d\tau=\mathbf{c}\,dt$, $m\mathbf{c}^2=m_0\mathbf{c}^2/(1-\beta^2)^{1/2}=m_0\mathbf{c}^2\gamma$, $u=-m_0\mathbf{c}^2/e$, $E=-E_x$, and for small displacements from the axis

$$E_z=-\frac{z}{2}\frac{\partial E_x}{\partial x},\qquad H_y=-\frac{z}{2}\frac{1}{\mathbf{c}}\frac{\partial E_x}{\partial t}.$$

Then

$$\frac{d}{d\tau}\left(\gamma u\frac{dz}{d\tau}\right)=-\frac{z}{2}\left(\frac{\partial E}{\partial x}+\beta\frac{\partial E}{\partial \tau}\right), \qquad \ldots\ldots(1)$$

so that the change in $p=\gamma u\,dz/d\tau$ in crossing the cavity is

$$p_2-p_1=-\tfrac{1}{2}\int\left(\frac{\partial E}{\partial x}+\beta\frac{\partial E}{\partial \tau}\right)z\,d\tau.$$

A first approximation is obtained by neglecting the variation of z and β in (1); if the field is nearly uniform across the gap then $\partial E/\partial x$ is concentrated at the ends and one finds

$$p_2-p_1=-\frac{z}{2\beta}(1-\beta^2)(E_1-E_2) \qquad \ldots\ldots(2)$$

where E_1, E_2 are the fields seen by the particle at the beginning and end of the gap. For phase stability E_1 exceeds E_2, so that the effect is a focusing one, but it becomes negligible as β approaches unity.

The second order term arising from the variation of β can be shown to be small compared with (2) for energies of more than a few MeV, but not so the term arising from the variation of z. This latter is again a focusing term, and comes about because the radial field at the entrance to the gap deflects the particle inwards; it is then nearer to the axis for the remainder of the gap, and the defocusing forces there are correspondingly less effective. We now consider only the extreme relativistic case so that $\beta = 1$, $dx = \mathbf{c}\,dt = d\tau$, and then, from eqn. (1),

$$\gamma u \frac{dz}{dx} - \gamma_1 u \left(\frac{dz}{dx}\right)_1 = -\tfrac{1}{2}\int_0^x z \frac{dE}{dx}\,dx\,.$$

But

$$\gamma u = \gamma_1 u + \int_0^x E\,dx,$$

so

$$\frac{dz}{dx} - \left(\frac{dz}{dx}\right)_1 = -\frac{1}{2\gamma_1 u}\left\{\int_0^x z\frac{dE}{dx}\,dx + 2\frac{dz}{dx}\int_0^x E\,dx\right\},$$

and if we use the fact that E is zero just before the gap

$$\frac{dz}{dx} - \left(\frac{dz}{dx}\right)_1 = -\frac{1}{2\gamma_1 u}\left\{zE - \int_0^x \frac{dz}{dx}E\,dx + 2\frac{dz}{dx}\int_0^x E\,dx\right\}. \quad \ldots\ldots(3)$$

Now if γ_1 is large, $dz/dx = (dz/dx)_1$, $z = z_1 + x(dz/dx)_1$. If we substitute these on the right-hand side of (3) we can calculate corrections of order γ_1^{-1}, and by substituting these in (2) again we obtain terms of order γ_1^{-2}, and so on systematically. Consider first the case $(dz/dx)_1 = 0$. The first approximation from (3) is $dz/dx = -z_1E/2\gamma_1 u$; at the end of the gap E becomes zero and so then does dz/dx; this we expected.

However,

$$z = z_1\left(1 - \frac{1}{2\gamma_1 u}\int_0^x E\,dx\right),$$

so that if l is the gap length

$$z_2 = z_1\left(1 - \frac{1}{2\gamma_1 u}\int_0^l E\,dx\right) = z_1\left(1 - \frac{1}{2\gamma_1}\right)$$

since for the synchronous particle the energy increase equals the rest energy. There is a fractional decrease in z of $1/2\gamma_1$: quite a large effect for low energy. Substituting $dz/dx = -z_1E/2\gamma_1 u$ back in (3) we have in the higher approximation

$$\left(\frac{dz}{dx}\right)_2 = -\frac{z_1}{(2\gamma_1 u)^2}\int_0^l E^2\,dx = -\frac{\kappa z_1}{4\gamma_1{}^2 l}$$

where κ is a dimensionless constant depending on the detailed field distribution; it is greater than unity, but usually only slightly so. To this order

$$p_2 = -\kappa u z_1/4\gamma_1 l. \quad \ldots\ldots(4)$$

Compare this with eqn. (2). If acceleration begins at the peak field and ends 37° later in phase, and if the field is uniform in the gap, then $E_1 - E_2 \sim E_1/5 \sim u/5l$. Thus (4) is greater than (2) by a factor of order 2γ; so already at a few MeV the effect of the variation of z is the more important, and this becomes increasingly so as the energy rises.

To complete the description of the action of the cavity on the vertical motion we must consider the case $z_1 = 0$, $(dz/dx)_1 \neq 0$. Of course $z_2 = l(dz/dx)_1$ in the first approximation, and from (3)

$$\left(\frac{dz}{dx}\right)_2 = \left(\frac{dz}{dx}\right)_1\left(1 - \frac{1}{2\gamma_1 u}\int_0^l E\,dx\right) = \left(\frac{dz}{dx}\right)_1\left(1 - \frac{1}{2\gamma_1}\right)$$

whence $p_2 = u\gamma_2\,(dz/dx)_2 = u(\gamma_1 + 1)(1 - 1/2\gamma_1)(dz/dx)_1 = p_1(1 + 1/2\gamma_1)$ to the required order of accuracy.

In general, then, the action of the cavity is represented by the matrix transformation

$$\begin{pmatrix} z_2 \\ p_2 \end{pmatrix} = \begin{pmatrix} 1-1/2\gamma_1 & l/\gamma_1 u \\ -\kappa u/4\gamma_1 l & 1+1/2\gamma_1 \end{pmatrix} \begin{pmatrix} z_1 \\ p_1 \end{pmatrix}. \qquad \ldots\ldots(5)$$

Now after leaving the cavity there are no vertical forces on the particle and the motion is one of drifting. The drift length is $\gamma_2\lambda - l = (\gamma_1+1)\lambda - l$, where λ is the wavelength of the radio-frequency field. The corresponding transformation is

$$\begin{pmatrix} z_1' \\ p_1' \end{pmatrix} = \begin{pmatrix} 1 & u^{-1}(\lambda - l/1+\gamma_1) \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_2 \\ p_2 \end{pmatrix}$$

and combining this with (5)

$$\begin{pmatrix} z_1' \\ p_1' \end{pmatrix} = \begin{pmatrix} 1-1/2\gamma_1-\kappa\lambda/4\gamma_1 l & u^{-1}\lambda(1+1/2\gamma_1) \\ -\kappa u/4\gamma_1 l & 1+1/2\gamma_1 \end{pmatrix} \begin{pmatrix} z_1 \\ p_1 \end{pmatrix}. \qquad \ldots\ldots(6)$$

The coefficients are correct to terms of the first order in γ^{-1}, and to this order the determinant of the transformation is unity, as it should be.

The transformation (6) is applied repeatedly, once for each revolution of the particle. This kind of problem has been discussed elsewhere (Le Couteur 1951, Bell 1953), and we will describe only the basic features of the solution. If we first neglect the variation of γ in (6) and denote the matrix by

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix}$$

it can be shown that z performs an oscillation $z = A\cos(\phi + \mu n)$ where n is the number of revolutions $(=\gamma-1)$ and $\cos\mu = (\alpha_{11}+\alpha_{22})/2$, which is in the present case $(1-\kappa\lambda/8\gamma l)$. Our approximation is good only when the last term is small, and then $\mu \sim (\kappa\lambda/4\gamma l)^{1/2}$ radians. If we put $\kappa\lambda/l \sim 10$ then μ varies from some 45° at $\gamma = 4$ (kinetic energy 1·5 MeV) to about 12° at $\gamma = 60$ (30 MeV). If now the α's vary slowly it can be shown (Bell 1953) that the amplitude of oscillation changes as $(\alpha_{12}/\sin\mu)^{1/2}$. In the present case α_{12} is nearly constant and $\sin\mu \sim \mu$, so $A \sim \mu^{-1/2}$. In going from $\gamma = 4$ to $\gamma = 60$ the amplitude about doubles and there are nearly three complete oscillations.

It must be pointed out in conclusion that any non-uniformity in the magnetic field can easily cause effects larger than those we have been calculating. This problem deserves detailed study, and it is important to consider also the oscillations in the meridian plane, which we have not touched on at all.

The Director of the Atomic Energy Research Establishment is thanked for permission to publish this communication.

REFERENCES

BELL, J. S., 1953, *Basic Algebra of the Strong Focusing System* (*Atomic Energy Research Establishment*, Report T/R 1114). (London : H.M. Stationery Office.)

HENDERSON, C., HEYMANN, F. F., and JENNINGS, R. E., 1953, *Proc. Phys. Soc.* B, **66,** 41.

LE COUTEUR, K. J., 1951, *Proc. Phys. Soc.* B, **64,** 1073.

SCHIFF, L. I., 1946, *Rev. Sci. Instrum.*, **17,** 6.

Magnetic Viscosity in Platinum Cobalt

By A. W. SIMPSON and R. H. TREDGOLD
Department of Physics, University of Nottingham

MS. received 13*th May* 1953

Detailed studies of the phenomena of magnetic viscosity have recently been made by various workers. In particular it has been shown that in the majority of ferromagnetic materials the magnetization, immediately after the application of a discontinuous change in the applied field, follows a law of the form

$$I = S \log t + \text{const.} \qquad \ldots\ldots (1)$$

where t is the time measured from the instant of the change in field and S is a parameter characteristic of the crystallographic and magnetic state of the material under consideration. Street and Woolley (1949) showed that this law can be explained in terms of an activation energy model if certain plausible assumptions are made, and Street, Woolley and Smith (1952) showed with Alnico II (Fe 54, Ni 18, Co 12, Al 10, Cu 6% by weight) that for a particular crystallographic state of the material the activation energy and applied field are related by

$$\delta E = \frac{q}{\boldsymbol{k}T}\,\delta H \qquad \ldots\ldots (2)$$

where δH is a small increment in effective field and δE is the corresponding mean change of activation energies; q is a parameter characteristic of the state of the material and evidently gives an indication of the average size of the region of magnetic material involved in the activation energy process. Barbier (1950) showed experimentally that there is an approximately linear relation between the reciprocal of the quantity denoted here by $q/\boldsymbol{k}T$ and the coercivity of a material. This relation appears to hold approximately over a wide range of coercivities from 1 to 1000 oersteds.

It is accordingly of interest to see the extent to which the behaviour of other materials, not yet studied in this connection, obeys Barbier's generalization. The present authors have therefore studied magnetic viscosity in an alloy containing equi-atomic proportions of platinum and cobalt and heat treated to have a coercivity of 1400 oersteds.

It is interesting to note that there is a similarity in the constitutions of the platinum–cobalt (Newkirk, Smoluchowski, Geisler and Martin 1951) and Alnico (Heidenreich and Nesbitt 1952) alloys in that in both cases it is possible to obtain a finely divided precipitated phase by quenching from a high temperature and subsequently annealing. It is thus not unlikely that the mechanisms responsible for the magnetic behaviours of the two series of alloys are very similar.

Street, Woolley and Smith (1952) also showed that for a material which can be described by an activation energy model

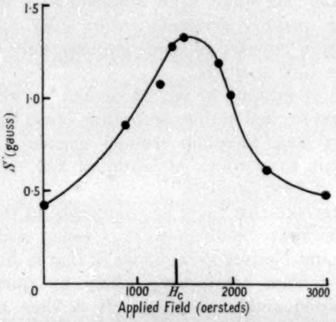
$$Sq/\boldsymbol{k}T = dI/dH \qquad \ldots\ldots (3)$$

where S and q are as defined in eqns. (1) and (2), and dI/dH is the irreversible susceptibility at the point at which the viscosity measurements are made. They also showed that this theoretical relation agrees well with experimental results in the case of Alnico II.

In the present work eqn. (3) is assumed to apply to platinum cobalt. This assumption appears to be justified because the behaviour of the material is reasonably well described by eqn. (1) and therefore can be discussed in terms of the activation energy model. The analogies between the metallurgies of platinum–cobalt and Alnico also appear to justify this step.

Since the magnetic fields necessary to make viscosity measurements on PtCo are well in excess of 1000 oersteds it was decided to use an electromagnet rather than a solenoid to apply the required field, a method which Street and Woolley have already used in this department and which will be described by them in the near future. The specimen is clamped between the pole pieces of the magnet and the potential proportional to the rate of change of magnetization induced in a coil wound on the specimen is measured. In the present work measurements were made with a galvanometer amplifier (Jones 1951) in order to obtain the requisite sensitivity and at the same time avoid the use of a large and cumbersome pick-up coil. This system was calibrated by the injection of a small known voltage into the primary galvanometer circuit. Owing to the high degree of stability required for the magnetic field, a valve stabilizer was used to maintain the magnet current constant.

In making measurements the specimen was brought into the cyclic state and the magnet current then adjusted to give a field approximately 80 oersteds less than the one at which it was proposed to make measurements. A discontinuous change was then made by means of the stabilizer circuit such that the final field was of the required value. Readings of the galvanometer deflection were then taken at fifteen-second intervals for thirty seconds to one hundred and twenty seconds after the instant of the discontinuous change.



Variation of S with applied field for CoPt.

It may be shown that, for a material whose behaviour is described by eqn. (1), the galvanometer deflections θ will be given approximately by

$$\theta = AS/t - B \qquad \ldots\ldots (4)$$

for large values of t. Here A and B are parameters characteristic of the apparatus and A may be calculated in terms of S.

The results obtained for an alloy having a composition corresponding to the formula CoPt and heat treated to have a coercivity of 1400 oersteds were in reasonable agreement with eqn. (4). Consequently, it is assumed that the behaviour of this alloy may be described approximately by eqn. (1).

The variation of the parameter S with applied field is shown in the figure. If one compares the results obtained for Alnico II (Street and Woolley 1949) fairly close agreement in the form of the curve and the relation of the maximum to the coercivity are found, notwithstanding the big difference in the coercivities of the two materials.

The value of the parameter $\boldsymbol{k}T/q$ found for this material from eqn. (3), using the measured value of dI/dH, is 8·5. This value is in excellent agreement with the linear relation between $\boldsymbol{k}T/q$ and H_c found for other materials by Barbier (1950).

ACKNOWLEDGMENTS

The authors wish to express their thanks to Professor L. F. Bates for the facilities of his laboratory and for the loan of the specimen, and to Dr. R. Street for helpful discussions.

REFERENCES

BARBIER, J. C., 1950, *C.R. Acad. Sci., Paris*, **230**, 1040.
HEIDENREICH, R. C., and NESBITT, E. A., 1952, *J. Appl. Phys.*, **23**, 352.
JONES, R. V., 1951, *Proc. Phys. Soc.* B, **64**, 469.
NEWKIRK, J. B., SMOLUCHOWSKI, R., GEISLER, A. H., and MARTIN, D. L., 1951, *J. Appl. Phys.*, **22**, 290.
STREET, R., and WOOLLEY, J. C., 1949, *Proc. Phys. Soc.* A, **62**, 562.
STREET, R., WOOLLEY, J. C., and SMITH, P. B., 1952, *Proc. Phys. Soc.* B, **65**, 679.

LETTERS TO THE EDITOR

Vapour Pressure Ratio of $^{12}C^{16}O$ and $^{13}C^{16}O$

Direct determinations have been made of the differences δp between the vapour pressures of a number of samples of carbon monoxide containing various mole fractions c of ^{13}C (respectively 72·79, 57·76, 44·75, 30·57 and 13·98 mole % of $^{13}C\,^{16}O$) and a 'standard' sample containing virtually no ^{13}C. Measurements have been made in the temperature range 68·2°K to 81·2°K, where the samples whose vapour pressures were being compared were in the liquid form, and in the range 61·6°K to 68·0°K, where they were solid. A few measurements were also made below 61·5°K (i.e. below the temperature of transition between the two solid forms).

A cryostat was constructed which could be maintained at any desired temperature in the range. The two samples whose vapour pressures were to be compared were condensed into two identical recesses in a well-insulated copper block A inside the cryostat, and the temperature T of the copper block was reduced to be slightly below that of the rest of the apparatus by pumping oxygen previously condensed in a third recess in A. The difference between the vapour pressures of the two samples was then measured on a differential bellows manometer of the type described by East and Kuhn (1946). The vapour pressure of the 'standard' sample was also measured absolutely by a mercury manometer, and the temperature T was deduced from the vapour pressure data of Clayton and Giauque (1932). The samples whose vapour pressures were to be compared were purified immediately prior to each series of experiments in a special high-efficiency column having a very small hold-up of material (<1 g). The isotopic composition of the actual samples used for the measurements was subsequently determined using a mass spectrometer.

It has been found that, within the limits of accuracy of the experiments, the vapour pressure of carbon monoxide is linearly related to the mole fraction of $^{13}C\,^{16}O$ present in the liquid or solid phase. Assuming that this is true even up to 100% $^{13}C\,^{16}O$, the vapour pressure ratio p_1/p_2 of $^{12}C\,^{16}O$ and $^{13}C\,^{16}O$ at T°K has been deduced from each measurement. The results for the liquid in the temperature range 68·2°K to 81·2°K may be represented by the formula

$$\frac{\Delta p}{p_1}=\frac{p_1-p_2}{p_1}\simeq\ln\left(\frac{p_1}{p_2}\right)=\frac{78{\cdot}2}{T^2}-\frac{0{\cdot}394}{T}.$$

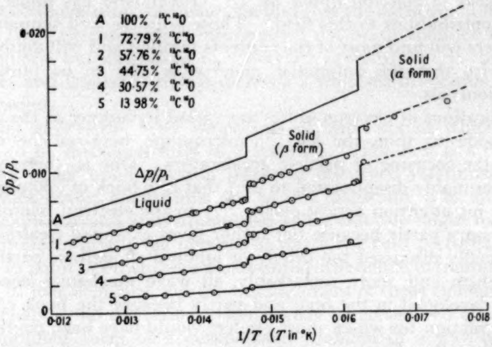
The value of $\Delta p/p_1$ varies from 0·0109 at 68·3°K to 0·0070 at 81·1°K, corresponding to values of Δp between about 1·25 and 5 mm Hg.

The data for the solid in the temperature range 61·6°K to 68·1°K may be represented by the formula

$$\ln\left(\frac{p_1}{p_2}\right)=\frac{82}{T^2}-\frac{0{\cdot}36}{T}.$$

At lower temperatures the measurements are rather inaccurate because the pressure differences being measured are so small; $\ln(p_1/p_2)$ varies from about 0·018 at 61·3°K to 0·021 at 57·2°K.

In the figure the measured values of $\delta p/p_1$ are plotted against $1/T$. The curves shown are calculated from the above formulae and the assumption that $\delta p = c\Delta p$. In no case does an experimental point for the liquid lie off the corresponding curve by more than 1% of the value of $\Delta p/p_1$ at that temperature, and in most cases the agreement is much better than this. For the β form of the solid the agreement is not quite so good, as the differences of pressure being measured were smaller, but the divergence of the experimental points from the curves is never more than 2%, and in most cases is only about 1% of the corresponding value of $\Delta p/p_1$.



Direct determinations have also been made of the differences between the melting and transition pressures of the same samples and from these measurements, together with the data on the differences of vapour pressures just above and just below the melting and transition regions, the melting and transition temperatures of the various samples have been calculated. In each case the temperature is found to be linearly dependent on the mole fraction of $^{13}C\,^{16}O$ present. The melting point of $^{13}C\,^{16}O$ is deduced to be $0{\cdot}061_5$°c higher than that of $^{12}C\,^{16}O$, and the transition temperature 0·112°c higher than that of $^{12}C\,^{16}O$.

Experiments are at present being made to determine the effect of isotopic abundance on the vapour pressures of a number of other substances. A more detailed account of the whole investigation will be published when these measurements have been completed.

I am indebted to Mr. D. Dance for making the mass spectrometric measurements, to Dr. H. London for his interest in the work, and to the Director of the Atomic Energy Research Establishment for permission to publish this letter.

Atomic Energy Research Establishment,
Harwell, Didcot, Berks.
20th May 1953.

T. F. JOHNS.

CLAYTON, J. O., and GIAUQUE, W. F., 1932, *J. Amer. Chem. Soc.*, **54,** 2610.
EAST, H. G., and KUHN, H., 1946, *J. Sci. Instrum.*, **23,** 185.

REVIEWS OF BOOKS

Grundlagen der Elektronenoptik, by WALTER GLASER. Pp. x+699. (Vienna: Springer-Verlag, 1952.) £10 4*s.* 6*d.*

Although there are several excellent comprehensive textbooks upon electron optics, until recently there has been barely a choice of monographs specifically devoted to a formal theoretical presentation of the subject. For this reason many will be glad to learn of the publication of a weighty volume upon the theoretical basis of electron optics by Dr. Glaser who has himself made such impressive contributions to this field. Those who are well acquainted with the author's papers will find most of the contents familiar and will doubtless note the author's advice that the unfamiliar material is also to be published in the appropriate journals.

The applications of electron optics are varied including, as the author points out, the cathode-ray tube, the electron microscope, beta-ray and mass spectrometers, and the focusing of particle accelerators. One is, therefore, surprised and not unreasonably disappointed to find that in a book of 700 pages the writer must restrict his attention almost exclusively to the electron microscope. This is made necessary partly because the author gives extended treatments of topics which are usually discussed too briefly or ignored altogether, partly because all electromagnetism and, more remarkably, all wave mechanics necessary to the argument is developed in the text, and partly because the book contains much detailed information for which most readers would have been content to turn to the periodicals. But, for these same reasons, those who are actively engaged in the design of electron microscopes and similar instruments should find this book of great value.

Almost half of the book is given to the first section upon '[Electromagnetic] Imaging Fields and Gaussian Dioptrics'. This contains the familiar dynamical treatment of electron lenses according to the paraxial approximation together with detailed discussion of analytical models, further expanded by the inclusion of such interesting, though perhaps somewhat academic, topics as the imaging condition regarded as an eigenvalue problem and the author's elegant theory of osculating cardinal points. One also finds a chapter given to experimental field measurement and path determination and to field calculation, and another given to the numerical calculation of paraxial rays but, although there is a numerical example of the relaxation method, the tendency is to treat methods of calculation formally; it is to be regretted that the reader is not given explicit guidance, clarified with examples, for computing the paraxial properties and aberrations of a practically determined lens field. It is also disappointing to find this distinguished author perpetuating the confusion which surrounds the short-lens formula for the focal length of electrostatic lenses; both the rival formulae are presented apparently without the author's realizing that they are incompatible. The most welcome feature of this section, as of the next, is undoubtedly the collecting together of the author's extensive calculations on analytical models of electron lenses, for these results are of great and proven value to designers.

The next quarter of the book is occupied by a section on the 'Theory of Geometrical Aberrations'. This begins unhappily with the republication of the author's original derivation of the formula for the refractive index of an electron-optical field although, as has been pointed out in the columns of this journal, this derivation is based on a complete misunderstanding of the principle of least action. It is also somewhat surprising to find that the author revives his original treatment of the third-order aberrations by means of the Seidel eikonal although this suffers from the disadvantages of being exceedingly tedious and of being valid only if the aperture is in field-free space. One cannot help feeling, at this stage, that had greater emphasis been laid upon the computation of the aberrations of experimentally determined lens fields, the writer would have endeavoured to improve upon the formulae which he first published some twenty years ago. In compensation, however, we find that the subsequent discussion of aberrations is most extensive and informative, not being confined to the usual geometrical figures, but extending to a detailed investigation of the associated caustic surfaces and isophotes which includes a welcome discussion of practical tests for aberrations.

This section ends with brief treatments of three topics which one could hardly expect to be discussed fully in a textbook. The first is a reproduction of Dr. Glaser's paper upon electron lenses of disturbed rotational symmetry. This work, which covers aberrations due to lens distortion, is as much as could be presented explicitly and is of undoubted value, but one might have expected the author to indicate briefly the manner in which the field perturbation is to be evaluated and to point out that there are essential differences in treatment between the distortion of lens elements and the misalignment of these elements.

The next chapter, on deflection systems, may be considered a model of how to treat a simplified form of a complicated problem in a way which brings out the salient features of the problem and indicates clearly the way to proceed with more complex forms. But the uninitiated may be distressed to find in use here a third method of aberration calculation; the method of variation of parameters has been used twice, the Seidel eikonal has been employed, and now there is introduced what is probably the most compact and flexible method, that of perturbation characteristic functions which the author first used in 1949. It would have made matters much clearer and simpler for the reader had one method, preferably the last, been used throughout or at least in the treatment of the third-order aberrations since this method is applicable for *any* aperture position.

There is a subtle fallacy in this chapter which it is worth while bringing to light. The perturbation procedure adopted by the author is valid for first-order calculations but it is suggested that extension to higher orders may be achieved by simple repetition of the first-order procedure, regarding the second term of the perturbing field as a perturbation of the unperturbed field together with its first perturbation. This seems at first sight to be plausible. However, in making a first-order calculation one commits, but neglects, errors of the second and higher orders: these errors cannot be evaluated by proceeding to the consideration of the next higher term of the perturbing field but only by re-inspection of the first-order calculation. The example chosen by the author fosters the misunderstanding since it is a system for which the first-order calculation happens to be exact so that there are no second-order errors to be picked up.

The last chapter of this section is, in strange contrast to the preceding chapter, a model of how not to simplify problems for didactic purposes since the specific topic treated bears little relation to the general problem to which it belongs. It reproduces a paper in which the author briefly discussed, by dynamical methods, the paraxial focusing of circular beams which encircle fields of rotational symmetry. This chapter gives no indication of how one should analyse the electromagnetic field for a more general problem; in reverting to dynamical methods it seems to imply that optical methods are no longer appropriate, which is untrue; and it provides the reader with no clue as to how he should proceed with aberration calculation.

The last quarter of the book is taken up by a section on ' Electron Optics and Wave Mechanics ', which the author opens by establishing wave mechanics from a comparison of Newton's and Huyghen's interpretations of Snell's law of refraction. One is also invited to consider such topics as the interpretation of wave mechanics in terms of a fictitious potential and several variations of Heisenberg's uncertainty relation before the wave theory of electron optics is embarked upon. But, once launched, this becomes the most interesting section of the book which gives a lucid presentation of diffraction theory according to the paraxial approximation and also taking account of aberrations. The section ends in fine style with a derivation of the Kirchhoff formula from the Schrödinger equation and an analysis of the resolving power of the electron microscope.

The text is fully illustrated and indexed, and there is an extensive bibliography in which the author lists and comments upon the publications of other workers. This affords some dry humour as, for instance, on the occasion when a 92-page paper by Picht and Himpan is annotated as a ' special case ' of a 16-page paper by the author.

The author's main object was to provide a didactic exposition of electron optics for the use of students and new-comers to the subject. There is no doubt that the book contains the essentials, and more than the essentials, of electron optics as applied to the electron microscope, but whether the subject has been made sufficiently tempting to induce a student to make such a very expensive purchase must be left for the student himself to decide. It is, in the opinion of this reviewer, as a reference work that this volume will be most highly valued : as such it is likely long to go unchallenged. P. A. STURROCK.

The Theory of Homogeneous Turbulence, by G. K. BATCHELOR Cambridge Monographs in Mechanics and Applied Mathematics. Pp. xi + 197. (Cambridge : University Press, 1953.) 25*s*.

Books on fluid dynamics generally tend to one or other of two extremes—either they are mathematical treatises with a few passing references to experimental evidence or they consist of résumés of laboratory experiments with a thin sprinkling of mathematics. Dr. Batchelor's timely account of recent advances in the more ' pure ' aspects of turbulence theory is remarkable if only for the fact that it pays considerable attention to laboratory measurements while remaining primarily a mathematical text.

One of the difficulties in formulating theories of turbulent motion is the lack of a truly mathematical definition of turbulence. Literary descriptions of the turbulent state are of little use and Dr. Batchelor inspires confidence by making his position clear from the first page. For him, a turbulent motion is one in which " the velocity takes random values which are not determined by the ostensible, or controlled, or ' macroscopic ' data of the flow, although we believe that the *average* properties of the motion are determined uniquely by the data ". (In 1951 the reviewer, when attending a conference in America, suggested that turbulent motion is one containing finite oscillations of velocity not specifically related to the boundary conditions, which is on much the same lines as Dr. Batchelor's definition.) The difficulty lies in the meaning to be attributed to the words ' random ', ' determined ' and ' macroscopic '. In principle, if the Navier–Stokes equations be accepted as universally true, the details of any motion can be found whenever the requisite initial and boundary conditions are known. In practice (apart from the technical difficulties of handling non-linear equations) it is impossible to be sure that all the relevant factors are specified, because all turbulence arises from instability and, initially, some of the factors may be below the threshold of observation. This is not to say that the velocity field is indeterminate in the strict sense of the word. It is made clear later in the book that what is implied is that the knowledge available from the ' macroscopic ' data is limited to a specification of the probability laws affecting the fluctuations. Turbulence theory then becomes the study of the statistical mechanics of a continuum, based on the belief that the population of fluctuations, as a whole, exhibits stable and predictable characteristics, even though the individual fluctuations defy prediction, or even exact description. This situation may arise because the laws which govern the continuum (the equations of viscous motion) are themselves statistical summaries of collision dynamics. Some more space might have been found in the book for a discussion of these points, which lie at the root of the problem.

Homogeneous turbulence is a special case in which, initially, an infinite fluid exhibits velocity fluctuations specified by probability laws which are independent of position, and the main problem is to trace the subsequent development of these laws in a continuous field subject to the equations of motion and of continuity. This is an ideal problem, for no known motion conforms entirely to these requirements, but the homogeneous field can be realized approximately in certain special arrangements (e.g. in the region far downstream of a regular lattice). Hence it is possible to use the mathematical theory to suggest quantities which can be measured and matched with the predictions of the theory. The object of this work is to investigate the basic properties of the turbulent field by considering a very limited case. There is no attempt to explore consequences which arise in technological work, such as the transport effects of the fluctuations.

The book opens with a general introduction and brief history of the problem. This is particularly well done. Chapters II and III, which deal with the mathematical representation of the turbulent field and the general kinematics of homogeneous turbulence, are not easy reading, especially for a newcomer. This is partly because of the real difficulties of the subject, but also because of the condensed and abstract form of the exposition and the lavish use of the double-suffix tensor notation. Undoubtedly at this stage the book suffers from an attempt to achieve too great generality too quickly and it would have been a better plan to introduce the subject by specific instances, such as the linear problems of Chapter IV. Chapter V on the general problem of decay also suffers in some respects from condensation, and the average reader will find Chapter VI which deals with the similitude theory of Kolmogoroff and others a welcome relief. Here, as might be expected from Dr. Batchelor's earlier papers, the treatment is excellent, the mathematics simple and straightforward and the physical concepts convincing and easily comprehended. The final chapters, dealing mainly with the energy-containing eddies, are well illustrated by experimental work, but occasionally one wishes for a more extended discussion of the laboratory results. In most of the examples it is immediately evident that the measurements lend considerable support to the theory but in others the comparison is not impressive at first sight. (Thus in fig. 4.2 the reader may be sceptical about the statement in the text that the experimental values are given " to a good approximation " by the theory.) The book concludes with a valuable bibliography, the compilation of which must have meant considerable labour.

Research workers will welcome this book as an indispensable account of the post-war activities of the Cambridge school founded by Sir Geoffrey Taylor. Those coming new to the subject may find the treatment austere and they would be well advised to acclimatize themselves by studying the later chapters first. Dr. Batchelor's main achievement undoubtedly is to have stated his problem explicitly and to have brought some much needed rigour into this difficult and diffuse subject.

The present volume is the first of a new series edited by Dr. Batchelor and Mr. Bondi. Other monographs are promised on shear turbulence, hydrodynamic instability, de Broglie waves and sound pulses, an interesting programme which has got off to an excellent start. The production of the book is a good example of the high standard which we have learned to expect from the Cambridge University Press. O. G. SUTTON.

Grundversuche der Physik in historischer Darstellung, by CARL RAMSAUER. Vol. I. *Von den Fallgesetzen bis zu den electrischen Wellen.* Pp. viii + 190. (Berlin, Göttingen, Heidelberg: Springer-Verlag, 1953.) DM. 19.80.

In the introduction to his excellent book Professor Ramsauer draws attention to one of those well-known facts that have a way of escaping attention, namely that the standard histories of physics by Poggendorff, by Heller and by Gerland contain no illustrations—and the illustrations in Rosenberger's history are so few that he might have added that work. Gerland and Traumüller's *Geschichte der physikalischen Experimentierkunst* is a valuable work which is copiously illustrated but stresses the apparatus rather than the advances made with its aid, and is little concerned with work after 1800. Professor Ramsauer, who is known to physicists for his fundamental investigations in electronics, including the effect known by his name, has set himself the task of describing certain well-chosen fundamental experiments in physics and the way in which they were carried out, writing throughout as an experimenter. This means, of course, that he has always gone to the original accounts and has read them with a full realization of the conditions of the times in question. The result is a book that will be of the greatest interest to all physicists with historical leanings. Probably any experimental physicist who picks up the book and starts reading will find that he has historical leanings.

The experiments range from those of Galileo on the law of fall to Hertz's investigations of electric waves, although one or two are included that fall outside the dates so determined. For instance, in discussing the history of the law of gravitation, the author takes together Tycho Brahe—somewhat before Galileo—Kepler and Newton. He admits that Newton's work in this field is scarcely experimental, but effectively defends its inclusion. Typical of his attitude is that Professor Ramsauer points out the fundamental importance of the accuracy which Tycho obtained with the instruments on which he lavished so much care.

These instruments had open, not telescopic, sights, but Kepler estimated that Tycho's error in angle did not exceed two minutes of arc: such was his confidence in this precision that a discrepancy of eight minutes between Tycho's observation and a calculation of the orbit of Mars which he, Kepler, had carried out on an early scheme of his was sufficient to lead him to abandon this scheme and turn to work that culminated in the discovery of the elliptic orbit. Professor Ramsauer is, perhaps, a little over-kind to Kepler's notions on the mechanics of the heavens—Kepler's ' magnetic ' force had no resemblance to a gravitational force, for he said quite plainly *non est attractoria sed promotoria*—but he rightly stresses his immense services.

Exceeding the limiting date at the other end is the account of Lebedew's experiments on the pressure of light, no doubt included as clinching Maxwell's theory of light and because the experimental elegance strongly appealed to the author. To give some notion of the scope of the book, the other experiments on light which are to be found here are those of Römer (perhaps hardly experimental), Fizeau and Foucault on the velocity of light, Newton's work on the decomposition of white light, Fresnel on interference, Malus on polarization and the chief optical work of Fraunhofer, Kirchhoff and Bunsen. The account of the experimental proofs of the kinetic theory of gases, dealing with the work of Robert Brown ('Brownian-movement Brown'), O. E. Meyer and Maxwell, is completed by a short description of the experiments which Stern carried out in 1920.

The careful discussion of the early experiments on the gas laws, going up to Regnault, disposes of Mariotte's claim to be included with Boyle. We are glad to see Andrews' fundamental experiments on the critical point included in the work on gases. The experiments on electricity and magnetism comprise, besides those that would at once occur to the reader, Rowland's work on the magnetic effect of a moving charge, a difficult and fundamental achievement. The book concludes with an account of Feddersen's investigations of the oscillatory discharge, a fascinating piece of work, and Hertz's discovery of electric waves, of which the author says that since Faraday's discovery of electromagnetic induction there has been no experimental discovery of such genius.

This book is written with a lifetime's knowledge of experimental physics and with an unabated enthusiasm for ingenuity of experimental method. It should meet with a warm welcome, not least wherever physics is taught. We look forward to a second volume which is announced, to cover the period from Röntgen rays to wave mechanics.

E. N. DA C. ANDRADE.

CORRIGENDUM

Electrical Conduction and Breakdown in Liquid Dielectrics, by D. W. GOODWIN and K. A. MACFADYEN (*Proc. Phys. Soc.* B, 1953, **66,** 85)..

The authors of the above paper state : " The dielectric constant of the liquid was unfortunately omitted from the analysis in § 5. It should appear as a multiplied factor on the left of eqn. (7) and throughout the subsequent equations. It should have been stated in § 5.2 that the field strength is assumed for simplicity to be modified by space-charge to only a small extent. In fig. 8 the labels should read 0, 100, 200, 300 μ. Equation (14) should read $F=bC_6/(2\sqrt{C_6}-1)$ instead of as shown, and the subsequent sentence should be deleted. These changes modify curve *a* in fig. 11, but, as is explained in § 5.3, the agreement is in any case partly fortuitous.

We regret these errors and apologize for any inconvenience caused to readers."

CONTENTS OF SECTION A

Dr. G. T. Wright. Fluoresence Characteristics of Mixed Organic Crystals . 777

Prof. D. R. Bates and Mr. G. Poots. Properties of the Hydrogen Molecular Ion—I: Quadrupole Transitions in the Ground Electronic State and Dipole Transitions of the Isotopic Ions 784

Dr. A. R. Edmonds. Nuclear Binding Energies and the *jj*-Coupling Shell Model . 793

Dr. J. G. Rutherglen and Dr. R. D. Smith. Excitation Curves of the Reactions ^{27}Al(p, α)^{24}Mg and ^{27}Al(p, γ)^{28}Si 800

Dr. L. R. B. Elton. Elastic Scattering of Electrons 806

Prof. L. F. Bates and Mr. A. Hart. A Comparison of the Powder Patterns on a Sample of Grain-Orientated Silicon–Iron with those obtained on a Single Crystal 813

Prof. L. F. Bates and Dr. G. W. Wilson. A Study of Bitter Figures on the (110) Plane of a Single Crystal of Nickel 819

Mr. G. D. Adam and Dr. K. J. Standley. Microwave Resonance Absorption in some Ferromagnetic Manganese Compounds 823

Dr. R. K. Laird and Dr. R. F. Barrow. The Ultra-Violet Band Spectrum of Carbon Monoselenide 836

Research Notes:

Dr. G. W. Wilson. A Proposed Structure for Certain Domain Configurations on a Single Crystal of Nickel 840

Mr. G. Manning and Mr. B. Singh. A Note on the ^{10}B(α, p)^{13}C Reaction . 842

Letters to the Editor:

Dr. G. K. White. Thermal Conductivity of Silver at Low Temperatures . 844

Dr. E. Breitenberger. On the Geometry of Angular Correlation Experiments 846

Dr. D. ter Haar. The Second Virial Coefficient near Absolute Zero . . 847

Mr. W. R. S. Garton. Observations of the 4050 Å Band Group in a King Furnace 848

Dr. B. Donovan and Dr. E. H. Sondheimer. Galvano-Magnetic Effects at High Frequencies 849

Reviews of Books 851

Contents of Section B 856

Surface Condition and Electrical Impedance in Rubber Friction

By A. SCHALLAMACH

British Rubber Producers' Research Association, Welwyn Garden City, Herts.

MS. received 27th April 1953

Abstract. It is shown experimentally, by employing electrically conductive rubber, that during the frictional gliding of rubber a thin layer in contact with the track undergoes a deformation which considerably increases the electrical impedance of this layer. In the case of previously abraded rubber the layer is permanent. If a d.c. potential is applied between sample and track during sliding, electrostatic attraction between sample and track is produced which increases the frictional force. The thickness of the layer, as estimated from the magnitude of the electrostatic force, is a few thousandths of a centimetre.

§ 1. INTRODUCTION

EVIDENCE has been advanced recently (Schallamach 1952 a) that the frictional force between rubber and a hard, smooth track is proportional to the true area of contact between rubber and track, the area depending on the deformation of the asperities on the surface of the sample under the influence of the normal load. It is to be expected that during sliding the asperities will undergo an additional deformation as a consequence of the tangential stress. Direct observation of the interface rubber-track is difficult, but it was thought that if the rubber could be made electrically conductive there should be detectable changes in the electric resistance between sample and track when the rubber is set into motion. Electrically conductive rubber is easily produced by incorporating carbon black in the compound, and the results obtained with such specimens are reported here. Great interest attaches to carbon-loaded rubbers for their own sake as practically all rubber tyres contain about 30% by weight of carbon black.

§ 2. EXPERIMENTAL

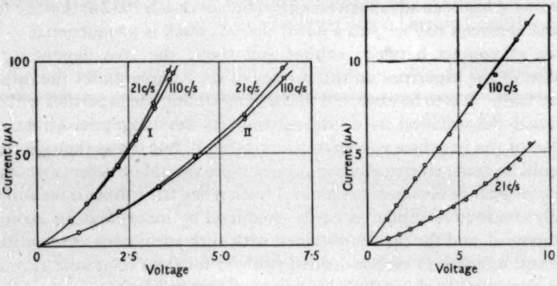
The equipment used for these experiments was the same as in the earlier investigation (Schallamach 1952 a, see also Roth, Driscoll and Holt 1942). Each sample, 3·17 mm thick and 4·7 cm^2 area, was bonded to a steel plate which served as one electrode. The other electrode was the track, made of chromium-plated brass, over which the sample slid at a velocity of 0·00216 cm sec^{-1}, and under a normal load of 3·96 kg. The track was cleaned according to a standard technique with acetone after each run. The composition of the samples was as follows: rubber 100, ZnO 5, S 2·5, Santocure (accelerator) 0·5, Philblack O (furnace black) 50·0, stearic acid 2·5, pine tar 4·5, phenyl-β-naphthylamine (antioxidant) 1·0. This recipe represents a typical tyre tread compound. The samples were cured at 140°c for 40 minutes, and their sliding surface was moulded against a ground glass plate.

The first measurements were carried out with a d.c. potential applied between sample and track, but it was found that the results were subject to time effects and that the samples did not obey Ohm's law. For this reason the experiments were continued with low-frequency alternating current. As will be shown later, Ohm's law was still not followed at these frequencies, but the deviations from linearity in the current–voltage curves were relatively unimportant, and measurements could be made within a narrow range of voltages which were kept constant in each individual run. The frequencies employed were 21 c/s and 110 c/s. These particular values were dictated, on the low-frequency side by availability, and on the high-frequency side by the endeavour to keep clear of the region of anomalous dielectric dispersion (Schallamach 1951). The alternating currents were measured by means of a valve voltmeter shunted by resistances of suitable values.

The ratio of potential difference to current will be referred to as impedance R, although this term may be slightly misleading in view of the non-linearity of the phenomena.

§ 3. IMPEDANCE CHANGES DURING SLIDING

When a new sample is put on the track and is loaded normally the impedance decreases and reaches a constant value after a few hours. For this reason the samples were left on the track for at least 12 hours before measurements commenced. The equilibrium current–voltage characteristics at the two frequencies are shown by curves I in fig. 1 (*a*), which demonstrate the extent to which Ohm's law is not obeyed by this material. The higher current at the lower frequency found under these conditions is most probably due to residual inductance in the leads and measuring equipment as the same effect was observed when the sample was replaced by an ohmic resistor. The corresponding difference in the nominal value of the impedance is small compared with effects observed during sliding.



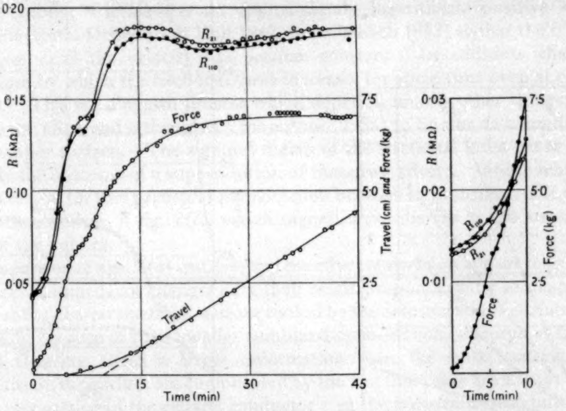
(*a*) (*b*)

Fig. 1. (*a*) Current–voltage characteristics of resting sample: I, when new; II, three days after completion of experiment. (*b*) Current–voltage characteristics of sliding sample after application of d.c. potential.

During sliding, the following measurements were made. The frictional force was read every 30 seconds, and at the same instants the impedances were determined, alternating measurements at 21 c/s and 110 c/s, so that the readings at each frequency were spaced at one minute intervals. The travel of the sample was determined at suitable times.

The results obtained with a new sample are given in fig. 2 (*a*), the impedance measurements having been made at a potential difference of 4·08 v. There is no material difference between the impedances at 21 c/s and 110 c/s, but they reverse their order during the initial stages of the experiment so that the low-frequency value becomes larger than the impedance at the higher frequency. The character of the impedance curves is very similar to that of the frictional force curve.

When the drive is switched on (time zero) the frictional force increases sharply, while the slack is taken up in various parts of the apparatus, mostly in the towing lines. There is a similar increase of the impedances during this period, and a point of inflection is found in all curves when static friction is overcome and the sample begins to move. The similarity in the time-dependence of the three quantities is also maintained in such details as the sigmoid range following shortly after the point of inflection. Finally, both frictional force and impedances become essentially constant, the latter after having passed through an ill-defined maximum.



(*a*) (*b*)

Fig. 2. (*a*) Impedances R, frictional force and travel of a new sample plotted against time. Applied voltage 4·08 v. (*b*) Impedances of, and shearing force on, rubber sandwiched between two bonded plates plotted against time. Applied voltage 1·61 v.

The sample does not immediately move with the prescribed constant velocity but starts at a slower rate and accelerates until the slack has been completely taken up. The maximum of the impedance curves appears to coincide with the region of the travel curve where the sample reaches constant velocity.

The changes of the impedance described above are only partially reversible. When the sample is stopped, the impedances decrease slowly but, as curves II of fig. 1 (*a*) show, they have still $2\frac{1}{2}$ times their original value after a lapse of 3 days.

Assuming that the effects shown by these experiments have their origin partly in the bulk of the rubber, and partly in the surface asperities, it is essential to assess the contribution to the effects by the deformation of the bulk material. For this purpose a sample was used which had approximately the same dimensions as the friction sample but was bonded between two steel plates, the lower one of which was fixed to the track and the upper one to the carriage usually holding the friction samples. The result of subjecting this sandwich to shearing stress is shown by fig. 2 (*b*). The force rises, of course, much faster than when the sample can move; the absolute values of the impedances and their relative increases at comparable values of the stress are considerably lower than those of the friction sample.

A simple variant of the experiment is to repeat the measurements after the rubber surface has been abraded. This was done by running the sample a few times over silicon carbide cloth (*John Bull*, No. 150), changing the direction of motion after each run in order to prevent the formation of an abrasion pattern (Schallamach 1952 b). As with a new sample, some time is required for the rubber to settle down after abrasion, and curves *a* in fig. 3 give the current–voltage characteristics after 24 hours. Abrasion affects the conductive behaviour of the sample profoundly. The impedance has increased one to two orders of magnitude, depending on the frequency. At 110 c/s the current is far larger than at 21 c/s and obeys Ohm's law within the limits of experimental error. The results obtained when the sample slides are given in fig. 4. The frictional force increases in the same manner as before, but the equilibrium value is about 15% lower than in the case of the new sample. From this may be deduced that the true area of contact has been reduced by a similar amount.

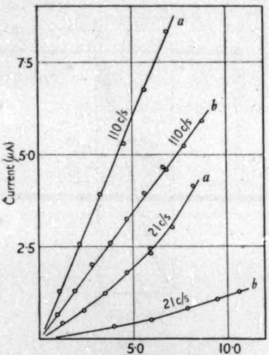


Fig. 3. Current–voltage curves of abraded sample: (*a*) resting sample, one day after abrasion; (*b*) during sliding, after application of d.c. potential.

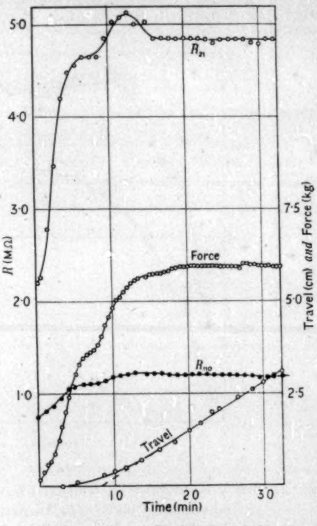


Fig. 4. Impedance *R*, frictional force and travel of abraded sample plotted against time. Applied voltages: at 21 c/s, 5·6 v; at 110 c/s, 4·0 v.

The trend of the impedances curves in fig. 4 is also the same as before but the relative increase at 110 c/s is very small compared with that at 21 c/s. The table gives equilibrium values of the relative increases of the impedance in the four instances described in this section.

Relative Increase of Impedance during Sliding

Frequency (c/s)	21	110
New sample	350%	330%
Abraded sample	230%	64%

It will be seen from this table that the effect is reduced by abrasion.

§ 4. DISCUSSION OF SLIDING EXPERIMENTS

After the sample has started to move in the experiments described above, the frictional force continues increasing. This is due to two causes. In the first instance rubber friction has an approximately logarithmic positive velocity coefficient (Roth, Driscoll and Holt 1942, Schallamach 1953) so that the frictional force rises until the velocity has become constant. In addition, there is a mechanism by which the frictional force increases for some time even at constant velocity. This is a sluggish process which depends, among other things, on the distance travelled, and is thought by the author (1953) to be due to a conditioning of the rubber surface. The sigmoid region of the frictional force curve is most probably the outcome of a superposition of these two effects. There may yet be another reason for this particular phenomenon because an indication of it is found in the stress curves of fig. 2 (*b*), which suggest a peculiarity in the stress–strain curve of the rubber.

Comparison of figs. 2 (*a*) and 2 (*b*) and the effect of abrasion appear to prove that the observed impedance changes have their origin preponderantly in the top layer of the rubber nearest the track, and are caused by the deformation of surface asperities which, because of their smaller combined cross section, undergo at the same applied shearing stress a larger deformation than the bulk material. The conditions on the surface are complicated by the fact that there are abrupt changes in the cross section of the electric conductor ; in the transition from bulk rubber to asperity, the total resistance is not equal to the sum of the geometrically calculated resistances, but there is an additional term called the spreading resistance (cf. Bowden and Tabor 1950, p. 25). Similar arguments apply to the impedance. The order of magnitude of the effect may be gauged by the simple case of a cylindrical asperity of equal diameter and height resting on a highly conductive track. The spreading resistance r at the junction of a small cylinder of radius a with an infinitely extended conductor is $r=\rho/4a$ where ρ is the specific resistivity of the material and amounts in the present instance to about 28% of the total resistance. If the asperity is strained during sliding, its cross section will most probably be reduced, and it will be seen that in this case the spreading resistance will increase. The conclusion drawn from these considerations is that the spreading resistance or impedance will enhance any effects due to deformation of the asperities.

The results obtained with the new sample will be dealt with in detail first. If this rubber were an ordinary, though poor, conductor of electricity the deformation of the asperities would by itself lead to an increased resistance by a 'strain gauge' effect. However, rubbers of the tyre tread type are not ordinary conductors, but their conductivity is highly strain sensitive (Bulgin 1945, Wack, Anthony and

Guth 1947). This behaviour is generally attributed to continuous, though open, carbon structure extending over large regions of the compound and readily breaking on straining the rubber (Mullins 1947). To estimate the amount of the strain of the asperities necessary to produce impedance changes of the order of those shown in fig. 2 (*a*) a cylindrical sample of 2·54 cm diameter and 1·27 cm height, bonded to steel plates, was extended and its impedance measured. The strain in this experiment is not well defined but may be considered as approximating to that of a simple extension. It was found that an extension of only 19% sufficed to increase the impedance by 350%, and a rough estimate shows that the same change of impedance could be brought about by a simple shear of about 35%, the angle of shear being slightly more than 19°.

During sliding the passage of alternating current through the sample ceases to be purely conductive as the impedance at 21 c/s exceeds that at 110 c/s. This behaviour indicates, of course, that a displacement current flows through the sample and that the total current is no longer the sole result of electronic and ionic conduction, or of the migration of other carriers of electricity.

With the abraded sample large local deformations occur on the abraded surface (Schallamach 1952 c) and the breakdown of the carbon strcuture brought about thereby is more pronounced than that brought about by the comparatively small strain occurring on the surface of a new sample during sliding. The top layer of the abraded rubber may be visualized as a dielectric containing well dispersed and immovable conductive carbon particles, which increase the dielectric constant of this layer, and discrete electricity carriers. The relative magnitude of the capacitive and conductive currents depends on the frequency, and at 110 c/s the contribution of the conductive current to the total current appears to be small because of the linearity of the current–voltage curve in fig. 3. The non-linearity of the corresponding curve at 21 c/s, like that of the curves in fig. 1, seems to show that the motion of the electricity carriers is not governed by Stokes' law.

As the conductive part of the current is confined to a path along the deformed asperities, but the capacitive current can pass from the solid back of the sample to the track by the dielectrically shortest way, it is not surprising to find that the impedance changes shown in fig. 4 should be greater at 21 c/s than at 110 c/s.

§ 5. THE EFFECT OF A D.C. POTENTIAL

If the top layer of high electric impedance postulated in the last section exists, the following effect may be predicted. When a d.c. potential is applied between sample and track, practically the whole voltage drop should take place across this top layer and thus give rise to a high electric field strength and to electrostatic attraction between sample and track, with a corresponding increase of the frictional force.

The results of such experiments are given in fig. 5 for a new sample, and in fig. 6 for an abraded sample. Relatively low voltages suffice for the production of substantial effects, a voltage of 240 v, for example, increasing the frictional force by nearly 50%.

On switching on the d.c. potential, the frictional force does not immediately rise to its final value. Within the first one or two minutes the frictional force rises quickly and then continues increasing at a lower rate for at least 30 minutes (cf. fig. 6) before it becomes constant.

These, and other, time effects shown in figs. 5 and 6 are almost certainly connected with the time dependence of the direct current flowing through the sample. The current decreases rapidly after being switched on; during the first five minutes it drops to about one third, and then becomes constant at one tenth, of the initial value. It is thought the declining current is due to the exhaustion of ions or other carriers of electricity in the top layer, thus leading to a higher resistance and higher field strength across the sample. That the passage of direct current produces a physical change in the sample is demonstrated by its subsequent behaviour towards alternating voltages. The changes brought about by the direct current are most pronounced in the new sample where, as fig. 1 (*b*) shows, the impedance is increased many times and has become frequency dependant. The effect is similar but not so pronounced in the abraded sample, as seen in fig. 3. These findings are compatible with the view that the number of electricity carriers has been reduced by the direct current.

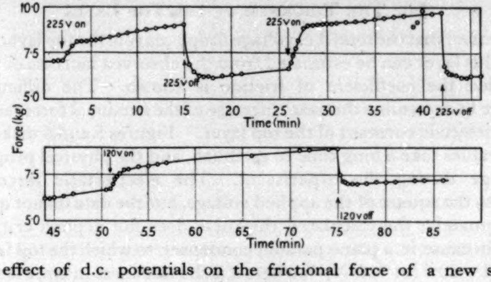


Fig. 5. The effect of d.c. potentials on the frictional force of a new sample plotted against time.

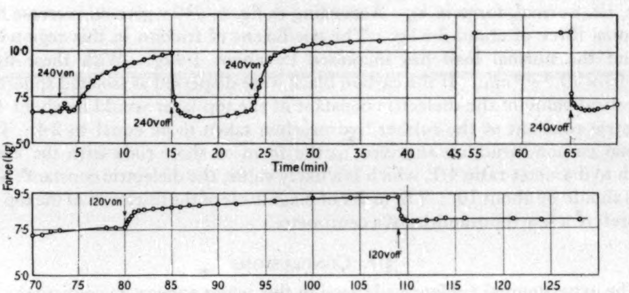


Fig. 6. The effect of d.c. potentials on the frictional force of an abraded sample plotted against time.

Electrostatic attraction can be shown by another, rather striking, method. Because of the positive velocity coefficient of rubber friction, rubber can be made to slide along a track under constant tangential stress (Schallamach 1953). If this is done with a sample of tyre tread rubber, application of a d.c. potential stops the sample instantaneously.

The electrostatic force does not, in principle, depend on the process of sliding, and it should be possible to demonstrate it with a resting sample of abraded rubber where, it has been pointed out, a permanent modified top layer exists. The difficulties in this experiment are the same as those encountered when the welded bridges between contiguous pieces of metal are to be evinced directly (Bowden and Tabor 1950); local contacts between sample and track must all be equally maintained while the normal load is removed. It has been found possible to make rubber adhere electrostatically to metal, but the results depended very much on the manipulative skill of the experimenter.

§ 6. THE THICKNESS OF THE TOP LAYER

If it is assumed that the total d.c. voltage drop occurs in the top layer, the effective thickness of this layer can be estimated from the observed increase of the frictional force, provided the coefficient of friction is known. The difficulties in this calculation are to determine the exact increase of the frictional force, and to find the value of the dielectric constant of the top layer. Figures 5 and 6 make it clear that equilibrium values take a long time to establish, and the physical properties of the sample change during the experiment. The electrostatic force should be proportional to the square of the applied voltage, but the data do not quite give this result. Fortunately, the calculated thickness does not depend critically on the other factors because in a plane parallel condenser, to which the top layer approximates, $d = 6{\cdot}71 \times 10^{-7} (\epsilon V^2 A/N)^{1/2}$, where d is the thickness in cm, ϵ is the dielectric constant, V the potential difference in volts, A the area in cm^2, and N the electrostatic, i.e. normal, force in kg. According to fig. 6, 240 v give an increase of the frictional force of about 3·6 kg. The coefficient of friction in this region is 1·9, so that the normal load has increased by about 1·9 kg. With these figures $d = 8{\cdot}03 \times 10^{-4} \epsilon^{1/2}$ cm. If the carbon black were dispersed as isolated spheres the theoretical value of the dielectric constant of the top layer would be about 4, the dielectric constant of the rubber hydrocarbon taken to be equal to 2·4. If the original carbon structure survived in the form of short rods with the average length to diameter ratio 4/1, which is a likely value, the dielectric constant (Sillars 1937) should be about 10. The order of magnitude of the thickness of the top layer is therefore a few thousandths of a centimetre.

§ 7. CONCLUSIONS

The experimental evidence adduced in this paper appears to confirm the view advanced in the Introduction that during the sliding of rubber a surface layer in contact with the track suffers considerable deformation. In the case of rubbers of the tyre tread type this top layer acquires physical properties different from those of the bulk material. In particular, its electrical resistance becomes very high and large electrostatic forces can be produced across it by means of moderate d.c. potential differences.

The top layer is to a certain extent permanently modified on abraded rubber. The existence of such a layer on the basis of other observations (Schallamach 1952 b) has been suggested at an earlier occasion.

ACKNOWLEDGMENT

This work forms part of a programme of research undertaken by the Board of the British Rubber Producers' Research Association.

References

BOWDEN, F. P., and TABOR, D., 1950, *The Friction and Lubrication of Solids* (Oxford: University Press).

BULGIN, D., 1945, *I.R.I. Trans.*, **21**, 188.

MULLINS, L., 1947, *J. Rub. Res.*, **16**, 275.

ROTH, E. L., DRISCOLL, R. L., and HOLT, W. L., 1942, *Bur. Stand. J. Res., Wash.*, **28**, 439.

SCHALLAMACH, A., 1951, *I.R.I. Trans.*, **27**, 40 ; 1952 *a*, *Proc. Phys. Soc.* B, **65**, 657 ; 1952 b, *I.R.I. Trans.*, **28**, 256 ; 1952 c, *J. Polym. Sci.*, **9**, 385 ; 1953, *Proc. Phys. Soc.* B, **66**, 386.

SILLARS, R. W., 1937, *J. Instn. Elect. Engrs.*, **80**, 378.

WACK, P. E., ANTHONY, R. L., and GUTH, E., 1947, *J. Appl. Phys.*, **18**, 456.

Thermal Effects at Point Contact Diodes

By P. M. TIPPLE and H. K. HENISCH
Physics Research Laboratory, University of Reading

Communicated by R. W. Ditchburn; *MS. received 29th May* 1953

Abstract. A thermoelectric method is described which enables the temperature of a point contact to be determined. Measurements have been made on a germanium diode at various positions on the voltage–current characteristic and it is found that voltage turnover occurs at a critical contact temperature, which is constant for a given specimen. The result also shows that temperature gradients in the neighbourhood of a hot point contact can increase the contact resistance. An interpretation of this observation is proposed.

§1. Introduction

It is well known that the properties of contacts between metals and semiconductors are generally sensitive to small temperature changes. The characteristics of such contacts are therefore influenced by the temperature rise which results from the dissipation of power in the contact region. A knowledge of the contact temperature, as distinct from the ambient temperature, is thus necessary if any comparison is to be made between calculated and measured voltage–current relations. Moreover, previous experiments by Hunter (1951), Billig (1952) and Bray (private communication) have shown that the peak turnover of a rectifying point contact diode is essentially a thermal phenomenon. For more detailed investigation of the conditions prevailing at or near turnover a method of measuring the actual contact temperature is required. Such a method is described in the present paper, and an account is given of some of the experimental results obtained using n-type germanium diodes, which have been made in the laboratory.

§2. Experimental Method

The principle of the present method is illustrated in fig. 1 (*a*). In position A of the switch the contact is subjected to a reverse load, which is maintained until thermal equilibrium at the contact is reached. When the current flow is interrupted (switch in position B), a thermo-e.m.f. will appear between the Y-plates of the oscillograph, since the temperature of the point contact, at which nearly all the power is dissipated, is higher than the base temperature. As the contact cools the thermo-e.m.f. decreases and its variation with time can be displayed on the oscillograph. The corresponding temperature difference can then be assessed from a previous calibration of thermo-e.m.f. as a function of temperature. The base temperature does not change appreciably in the course of electrical loading and could be conveniently measured by means of a thermocouple (not shown). The thermoelectric calibration experiment was carried out using large soldered electrodes in which thermocouple wires were embedded.

A simple test was made to confirm that the calibration curve so obtained does not differ materially from one obtained by means of tungsten point contacts.

In the arrangement actually used the vibrating relay and the battery are replaced by an electronic switching unit and a pulse generator. This is shown in fig. 1 (*b*). The circuits used are of conventional design. There is a delay between the cessation of the heating pulse and the instant at which the thermo-e.m.f. can first be measured; this interval depends on the circuit parameters. The contact temperature at zero time, i.e. immediately after the cessation of the heating pulse, must be obtained by an extrapolation which is performed graphically. In this way the contact temperature corresponding to steady electrical loads can be determined for various voltages applied to the contact and also for different ambient (base) temperatures. By way of example, fig. 2 (*a*) shows the relation between the power dissipated and the corresponding temperature rise at the point contact. The straight line obtained refers to results for a variety of ambient temperatures between 0°c and 100°c and extends over the whole range of powers which can be dissipated without causing permanent damage to the diode. The linearity of this relation is regarded as a valuable internal consistency test for the correctness of the temperatures measured, though it is not by itself an absolute confirmation. In fig. 2 (*b*) the decay of point temperature is shown as a function of time after cessation of the heating pulse. As might be expected, the relation is not characterized by a

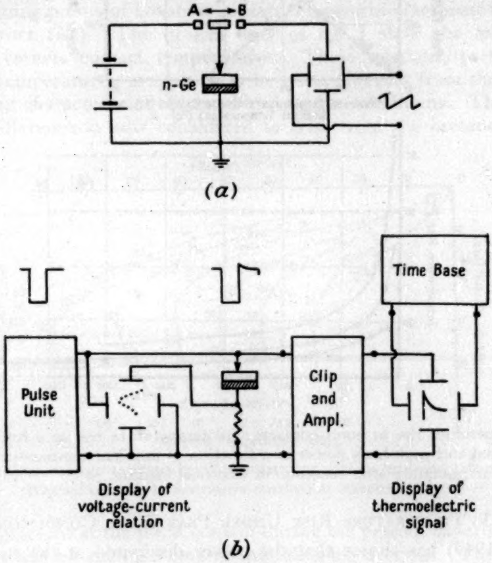

A
B
n-Ge
(*a*)
Time Base
Pulse Unit
Clip and Ampl.
Display of voltage-current relation
Display of thermoelectric signal
(*b*)

Fig. 1. Thermoelectric method of determining contact temperature : (*a*) schematic circuit; (*b*) experimental arrangement.

single time-constant. The times involved in the decay are in good agreement with those calculated by Burgess (private communication) for typical cases and also with those which may be roughly inferred from the results obtained by Billig (1952).

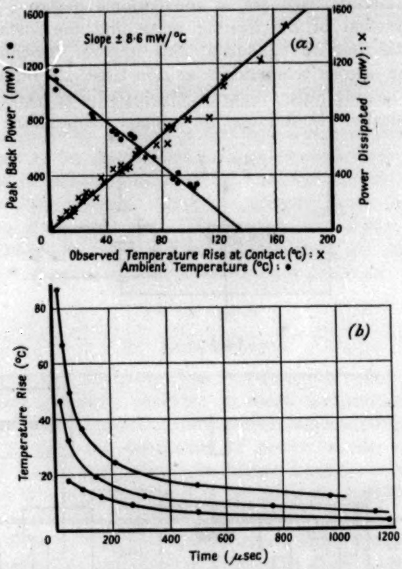


Fig. 2. Temperature rise at point contacts : (*a*) temperature rise as a function of power dissipated and peak back power as a function of ambient temperature; (*b*) decay of contact temperature after cessation of electrical loading.

§3. TEMPERATURE RISE UNDER PEAK BACK CONDITIONS

Benzer (1949) has shown that the power dissipated at the turnover point (the peak back power) decreases linearly with increasing ambient temperatures over an important part of the temperature range. Deviations from this relation are observed at high temperatures, e.g. above 100°c, but under those conditions the turnover point is no longer well defined. A plot of peak back power against ambient temperature is shown in fig. 2(*a*). According to Benzer's suggestion this relation implies that the turnover point corresponds to a definite point contact temperature which, in this case, is about 135°c as given by the intercept. This interpretation is supported by the present results. A comparison of the two lines in fig. 2(*a*) shows that their slopes are numerically identical within the limits of experimental error, as expected on the Benzer hypothesis. Moreover, the point contact temperature corresponding to voltage turnover can be independently determined by direct measurement. The value thus found is in good agreement with the value inferred from the intercept in fig. 2(*a*). Only at higher ambient temperatures (above 80°c) are slight systematic deviations observed. The germanium used for these experiments has a room temperature resistivity of 6 ohm cm and its maximum resistivity occurs at 40°c. A high proportion of intrinsically activated carriers are therefore present at a temperature of 135°c.

§4. Voltage–Current Relations for Constant Contact Temperature

Figure 3 gives a series of static characteristics observed at different ambient temperatures. At various points of these characteristics the contact temperature rise was also determined, and the corresponding values are marked. In this way it is clearly possible to evaluate a form of isothermal voltage–current relation by joining points of constant contact temperature (ambient temperature plus temperature rise). The broken lines of fig. 3 show the relations thus obtained for various contact temperatures. These relations, particularly for high contact temperatures, are found to be quite different from the isothermal voltage–current characteristics observed under pulse conditions. This originally unexpected difference is now considered to arise from the presence of a high temperature gradient at the point contact during the present experiments. This interpretation will be substantiated in the following sections. For the present it will be noted that the various points on each broken line do in fact correspond to different temperature gradients between the hot rectifying contact and the cold base, though the contact temperature itself is constant.

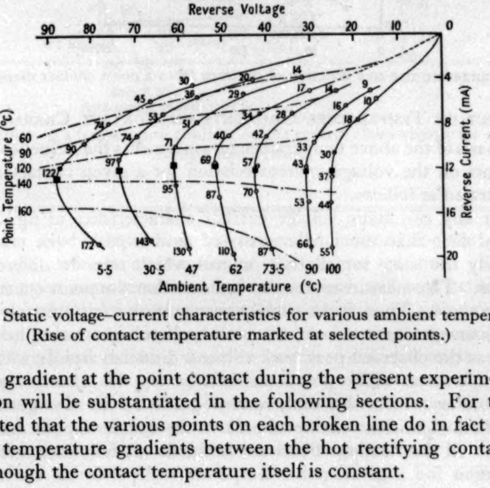


Fig. 3. Static voltage–current characteristics for various ambient temperatures. (Rise of contact temperature marked at selected points.)

§5. Temperature Gradient near a Rectifying Contact

A greatly simplified analysis, based on a hemispherical radius of contact and a constant heat conductivity, shows that the temperature gradient at any point in the germanium corresponding to a radius r is given by $(dT/dr)_r = -(T_r - T_\infty)/r$. Thus, at the contact itself, typical values may be $r_0 = 8 \times 10^{-4}$ cm and $T_r - T_\infty = 80$°c, which implies a temperature gradient of 10^5 deg c cm^{-1}. The experimental method described above also lends itself to the determination of such temperature gradient. It is merely necessary to place two point contacts in close proximity, one being connected to the pulse generator and the other to the clipping and amplifying equipment shown in fig. 1 (*b*). In this way the second point contact can be used to explore the temperature distribution in the neighbourhood of the first. Some disturbing influence due to the free surface and to the second (cold) point itself must be expected. Nevertheless, a relation of the above form is obtained, as shown in fig. 4. The temperature gradient at the contact itself, assessed from the graph for a total temperature difference of 80°c, is of the same order. It is clear from this that the temperature gradients which exist in the neighbourhood of a point contact under electrical load may be very large.

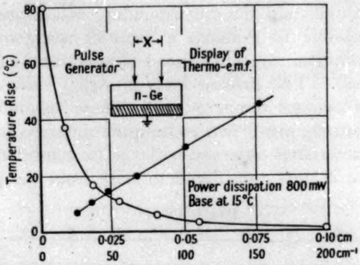


Fig. 4. Temperature rise as a function of distance from a point contact dissipating power.

§6. Effect of Temperature Gradients on Contact Characteristics

On the basis of the above temperature measurements the influence of temperature gradients on the voltage–current relation for a given contact temperature can be illustrated as follows.

Consider any two static voltage–current characteristics of fig. 3. It is an experimental fact that the currents passed under peak back conditions are approximately the same for all these curves which refer to different ambient temperatures. The measurements also indicate that the point contact temperatures are the same. Thus, if the contact temperature itself uniquely determines the voltage–current characteristic, the peak back voltages would have to be the same, whereas the observed peak back voltages diminish rapidly with increasing ambient temperature. The points marked by squares clearly correspond to conditions under which the temperature gradients in the germanium are different. Such conditions can never arise in the course of pulse measurements during which all the heating effects are avoided. It follows that the voltage–current relation for a given point temperature depends on the temperature gradient in the neighbourhood.

An interpretation which involves a critical contact temperature at turnover inevitably implies such a dependence.

The consistency of the present results and interpretations can be checked by the following experiment. The full lines in fig. 5 show two normal voltage–current characteristics for ambient temperatures of 30°c and 50°c respectively. The contact temperatures are also marked at selected points. The rectifying contact can be heated in two ways: by the power dissipated in the rectifier or, externally, by means of a subsidiary heater applied to the whisker as shown. In the confirmatory experiment the base temperature is kept constant at 30°c while the contact temperature (measured thermoelectrically by a potentiometer), in the absence of any applied voltage, is increased to 50°c by external heating. The voltage–current relation then obtained is represented by the broken line of fig. 5. This line can now be compared with the normal curve previously obtained for an ambient temperature of 50°c in the absence of external heating. In the two cases the starting temperature of the contact is the same, and the conditions for zero voltage are affected only by the presence of a standing temperature gradient due to external heating. It will be seen that the presence of this gradient has diminished the contact conductance everywhere.

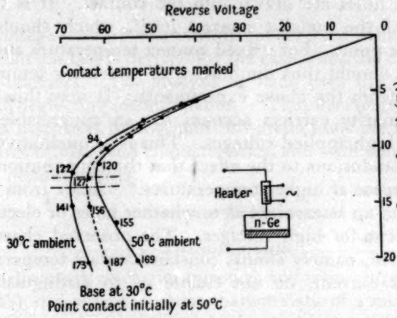


Fig. 5. Effect of a temperature gradient on the reverse characteristic for a given contact temperature.

§7. Discussion

The present experiments were not originally designed for the purpose of investigating the detailed mechanism of the reverse current. The results nevertheless suggest some tentative conclusions which will now be discussed.

When the *isothermal* voltage–current characteristics of point contacts are investigated by pulse methods they are found to be almost linear (except for very high voltages) and radiating from the origin at different angles for various ambient temperatures. This shows that heating is a necessary condition of the turnover phenomenon and the appearance of negative resistance. This is in agreement with the present results. On the other hand, it has already been noted that the lines joining constant contact temperatures in fig. 3 differ considerably from those obtained under pulse conditions. For a given contact temperature the two sets of experimental conditions differ only by the presence of temperature *gradients* at the contact during the thermoelectric measurements. These gradients are, of course, absent during pulse measurements. It must therefore be concluded that the presence of such temperature gradients has an important effect on the shape of the voltage–current relation for a given contact temperature. The effect is to diminish the point conductance, at any rate in the region before turnover.

There are, in principle, two mechanisms which could account for these observations. If all, or an appreciable fraction, of the reverse current under the conditions investigated is carried by electrons emitted from the metal, the above results would require that the barrier height should increase in the presence of the temperature gradients. This hypothesis demands a complicated model which has not yet been clearly formulated. A simpler interpretation can be obtained if it is assumed that an appreciable fraction of the reverse current under the present conditions is carried by minority carriers from the germanium. Though satisfactory numerical agreement between theory and experiment has not yet been obtained, the flow of minority carriers is believed to be essentially determined by the replacement mechanism (Bardeen and Brattain 1949). The rate of replacement should depend quite sensitively on the temperature of the region from which holes are drawn into the contact. It is this temperature, rather than that of the metallic contact itself, which should determine the current through the point. For a fixed contact temperature and applied voltage the reverse current should thus diminish with increasing temperature gradient, as is also inferred from the above experiments. It may thus be provisionally concluded that minority carriers account for an appreciable fraction of the reverse current for high applied voltages. This is in qualitative agreement with Bardeen's (1950) conclusions to the effect that the contribution of the minority carriers should increase at higher temperatures.* Apart from this, the present results do not lead to an assessment as to whether holes or electrons *predominate* in the reverse current for high voltages. The observed characteristics of the turnover phenomenon, namely almost constant contact temperature and almost constant peak back current, do not enable us to distinguish between these hypotheses.

ACKNOWLEDGMENTS

The authors wish to thank Professor R. W. Ditchburn for placing facilities at their disposal, Dr. P. C. Banbury and Mr. R. E. Burgess for helpful discussions, the Royal Naval Scientific Service for their support, and the Admiralty for permission to publish this paper.

REFERENCES

BANBURY, P. C., 1953, *Proc. Phys. Soc.* B, **66,** 50.
BARDEEN, J., 1950, *Bell. Syst. Tech. J.*, **29,** 469.
BARDEEN, J., and BRATTAIN, W. H., 1949, *Phys. Rev.*, **75,** 1208.
BENZER, S., 1949, *J. Appl. Phys.*, **20,** 804.
BILLIG, E., 1952, *Phys. Rev.*, **87,** 1060.
HUNTER, L. P., 1951, *Phys. Rev.*, **81,** 151.

---

* The interpretation is also consistent with Banbury's (1953) observation of carrier extraction effects.

Theory of the Forward Characteristic of Injecting Point Contacts*

By P. C. BANBURY
Physics Research Laboratory, University of Reading

Communicated by R. W. Ditchburn; *MS. received 8th May* 1953

Abstract. The voltage–current relationship for forward voltages across a metal–semiconductor point contact is derived for the case when the current is carried by injected minority carriers. The analysis is made with the assumption of unit injection ratio, and neglecting recombination. The effect of departures from these approximations is discussed briefly, and the predictions are compared with the well-known features of observed characteristics, and with some photoconductive and photovoltaic measurements.

§ 1. Introduction

PRIOR to the discovery of carrier injection and transistor action (Bardeen and Brattain 1948), the voltage–current characteristic of a metal point contact on n-type germanium was generally compared with the predictions of the diode theory of Bethe (1942), derived from the barrier model, in which the contribution of minority carriers (holes in n-type material) to the current was neglected. The predicted form of the voltage–current relationship for the barrier, with the usual approximations, may be written :

$$i=\frac{4\pi \boldsymbol{mek}^2T^2}{\boldsymbol{h}^3}\exp\left(-\frac{\phi}{\boldsymbol{k}T}\right)\left[\exp\left(\frac{\boldsymbol{e}V_{\mathrm{c}}}{\boldsymbol{k}T}\right)-1\right] \qquad \ldots\ldots(1)$$

where ϕ is the potential difference between the Fermi level in the metal and the lowest level of the conduction band in the semiconductor at the interface (see fig. 2), V_{c} is the applied voltage *across the barrier* in the semiconductor, such that V_{c} is positive in the low resistance direction, and the remaining symbols have their usual meaning. This equation or its equivalent forms are well known; some of the discrepancies which exist between predictions from this theory and measurements have been discussed, for example, by Torrey and Whitmer (1948). Two main features of disagreement for forward applied voltages are of interest here. It was to be expected on this model that for forward voltages small compared with the equilibrium barrier height the main contribution to the resistance of the system, comprising contact barrier and bulk materials, should arise from the barrier itself. From eqn. (1) we have, when $\boldsymbol{e}V_{\mathrm{c}} \gg \boldsymbol{k}T$,

$$\log i \simeq \frac{\boldsymbol{e}V_{\mathrm{c}}}{\boldsymbol{k}T} + \text{constant}.$$

Hence a straight line relationship of slope $\boldsymbol{e}/\boldsymbol{k}T$ was expected between $\log i$ and V_{c}. For larger values of external applied voltage it was usual to make an estimate of the voltage V_{c} across the barrier by applying a correction for the existence between barrier and base electrode of a spreading resistance R_s, assumed constant. Then for larger applied voltages also, with this correction, the same straight line relationship was expected. In practice it was found that no single value of R_s gave a straight line relationship, and further, that even in the region where the series resistance was expected to be relatively small a slope was obtained which was less than $\boldsymbol{e}/\boldsymbol{k}T$ by a variable factor β of the order of $\frac{1}{2}$.

---

* This paper is a summary of a part of the thesis submitted by the author for the degree of Ph.D. in the University of Reading (May 1952).

Since the discovery of carrier injection, the non-ohmic character of R_s has been qualitatively understood. No general discussion of the forward characteristic of point contacts has, however, been given. The discovery of injection did not involve the abandonment of the barrier model, but the diode treatment is clearly inadequate for discussing the minority carrier current, since the flow does not in this case occur over a potential barrier, and so is not limited to a small exchange between two systems in internal equilibrium. For this reason the assumption that the resistance of the system must be located in the barrier has lost its validity even for small applied voltages. The whole question of the forward characteristic requires reconsideration. Some discussion of the forward current in the presence of injection has been given by Bardeen and Brattain (1949), and this treatment will be extended here.

§2. THEORY OF THE FORWARD CHARACTERISTIC

The aim of the following analysis is to examine, under certain simplifying assumptions, the shape of the forward characteristic to be expected in the presence of injection. The argument as stated applies to n-type material. The following initial assumptions are made: (*a*) the current is carried entirely by holes, i.e. the injection ratio is unity; (*b*) the current flows radially from a hemispherical contact; (*c*) recombination of injected carriers may be neglected.

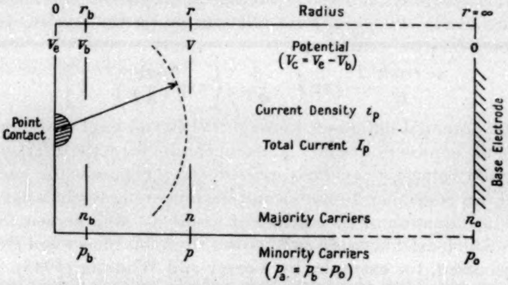


Fig. 1. Diagram illustrating symbols used in analysis.

The errors introduced by the last of these assumptions are expected to be small in a point contact system, since the spreading of the current in the semiconductor results in a dilution of the injected carrier concentration with increasing r even in the absence of decay.

The meaning of the symbols used in the analysis is illustrated in fig. 1. The distance r is measured from the centre of the hemispherical contact, and n, p and V represent the electron concentration, hole concentration and potential respectively at distance r.

As a result of assumptions (*a*) and (*c*) the electron current is everywhere zero, and hence

$$-n \operatorname{grad} V = -\frac{\boldsymbol{k}T}{\boldsymbol{e}} \operatorname{grad} n. \qquad \ldots\ldots(2)$$

Integrating and putting $V=0$ at $n=n_0$, we have

$$V = \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log\left(\frac{n}{n_0}\right). \qquad \ldots\ldots(3)$$

The value of n at any point in the semiconductor outside the barrier may be related to p by imposing the condition of electrical neutrality. This condition requires that any addition of holes should be balanced by an equal number of added free electrons (complete ionization of impurity centres being assumed), giving $n=n_0+p-p_0$. Substituting in (2), we have

$$\operatorname{grad} V = [\boldsymbol{k}T/\boldsymbol{e}(p+n_0-p_0)] \operatorname{grad} p. \qquad \ldots\ldots(4)$$

The general expression for hole current density i_{p} is given by

$$i_{\mathrm{p}} = -p\boldsymbol{e}\mu_{\mathrm{p}} \operatorname{grad} V - \boldsymbol{k}T\mu_{\mathrm{p}} \operatorname{grad} p. \qquad \ldots\ldots(5)$$

Combining (4) and (5), it follows that

$$i_{\mathrm{p}} = -\boldsymbol{k}T\mu_{\mathrm{p}}\left(\frac{p}{p+n_0-p_0}+1\right) \operatorname{grad} p. \qquad \ldots\ldots(6)$$

It may be seen here that the condition of zero electron current implies that the contribution to the hole current arising from the diffusion process (represented by the second term in the right-hand side of eqn. (6)) is everywhere greater than, or equal to, that arising from field effects (represented by the first term). Hence a discussion in terms of conductivity $\sigma = ne\mu_{\mathrm{n}} + pe\mu_{\mathrm{p}}$ cannot be adequate in these circumstances.

Proceeding, eqn. (6) may be re-written

$$\frac{I_{\mathrm{p}}}{2\pi r^2 \mu_{\mathrm{p}} \boldsymbol{k}T}\, dr = -\left(2 - \frac{n_0-p_0}{p+n_0-p_0}\right) dp$$

where I_{p} is the total hole current through the contact. Integrating and inserting the condition $p=p_0$ at $r=\infty$, we have

$$\frac{I_{\mathrm{p}}}{2\pi r \mu_{\mathrm{p}} \boldsymbol{k}T} = 2(p-p_0) - (n_0-p_0) \log\frac{(p+n_0-p_0)}{n_0}. \qquad \ldots\ldots(7)$$

It is now convenient to consider two ranges of values of the hole concentration p.

(*a*) $p \ll n_0$. In this range (7) may be written

$$p - p_0 \simeq \frac{I_{\mathrm{p}}}{2\pi r \mu_{\mathrm{p}} \boldsymbol{k}T}. \qquad \ldots\ldots(8)$$

(*b*) $p \gg n_0$. In this range (7) reduces to

$$p \simeq \frac{I_{\mathrm{p}}}{4\pi r \mu_{\mathrm{p}} \boldsymbol{k}T}. \qquad \ldots\ldots(9)$$

The second of these two expressions is that obtained by Bardeen and Brattain (1949), giving an approximation for the hole concentration for small r or large I_{p}.

Now let V_{e} be the total external voltage applied between the metal and the semiconductor where r is large. Let the subscript b signify values at the boundary between the barrier and the bulk semiconductor. Then, since $V=0$ at the base contact, V_b represents the voltage across the bulk semiconductor. From eqn. (3) it follows that

$$V_b = \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log \frac{n_b}{n_0} = \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log \left(\frac{n_0 + p_b - p_0}{n_0} \right). \qquad \ldots\ldots (10)$$

Now, in equilibrium, when $V_e = 0$, $p_b = p_0$. Further, since the barrier is of thickness only a few times the mean free path for holes, it will be assumed that the dependence of p_b on the voltage across the barrier is given by the equilibrium energy distribution of holes at the contact. The resistive drop due to hole flow across the barrier is thus neglected. The applied voltage across the barrier is $V_e - V_b$, and so we have

$$p_b = p_0 \exp \frac{\boldsymbol{e}}{\boldsymbol{k}T} (V_e - V_b), \qquad \ldots\ldots (11)$$

i.e.

$$V_e - V_b = \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log \frac{p_b}{p_0}. \qquad \ldots\ldots (12)$$

It follows immediately from (11) and (8) or (9) that the current flowing is related to the voltage across the barrier by an expression of the form

$$I = A \exp \{ \boldsymbol{e}(V_e - V_b)/\boldsymbol{k}T \} \qquad \ldots\ldots (13)$$

in the range $\boldsymbol{e}(V_e - V_b) \gg \boldsymbol{k}T$, where the term A varies within a factor of 2 over the range of values of I, and so changes only slowly in comparison with the exponential term for varying V. Thus, neglecting the small changes in A, this analysis also requires that a plot of log I against $V_e - V_b$ should give a line of slope $\boldsymbol{e}/\boldsymbol{k}T$. As yet, however, no explicit value of V_b, the voltage across the 'spreading resistance', has been obtained in terms of known quantities. This could be done by combining eqns. (10) and (8) or (9). An alternative treatment will be adopted here, however.

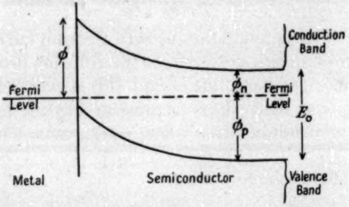


Fig. 2. Schematic diagram of electron energies at contact barrier, showing notation used in text.

We have, from (10) and (12),

$$V_e = \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log \left(\frac{p_b}{p_0} \cdot \frac{n_0 + p_b - p_0}{n_0} \right).$$

Introducing a new symbol for the added hole concentration at the barrier, $p_a = p_b - p_0$, this gives

$$V_e = \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log \left(1 + \frac{p_a}{p_0} + \frac{p_a}{n_0} + \frac{p_a^2}{p_0 n_0} \right). \qquad \ldots\ldots (14)$$

We can now conveniently deal separately with the two cases (*a*) and (*b*) distinguished previously.

(*a*) $p_b \ll n_0$ *and so* $p_a \ll n_0$.

Here, for very small p_a, (14) reduces to

$$V_e \simeq \frac{\boldsymbol{k}T}{\boldsymbol{e}} \log \left(1 + \frac{p_a}{p_c} \right). \qquad \ldots\ldots (15)$$

As p_a increases, while still satisfying (*a*), the fourth term in the logarithm on the right-hand side of (14) will become larger than unity, but in these circumstances the term p_a/p_0 is dominant, and the expression may be approximated by neglecting all other terms. Therefore the above approximation will be used. Substitution from (8) leads to

$$I_p = 2\pi r_b \mu_p \boldsymbol{k} T p_0 \left[\exp\left(\frac{\boldsymbol{e} V_e}{\boldsymbol{k} T}\right) - 1 \right].$$

p_0 may be related to the equilibrium Fermi level by the expression

$$p_0 = 2\left(\frac{2\pi \boldsymbol{m} \boldsymbol{k} T}{\boldsymbol{h}^2}\right)^{3/2} \exp\left(-\frac{\phi_p}{\boldsymbol{k} T}\right) \qquad \ldots\ldots(16)$$

where ϕ_p is the energy difference between the Fermi level and the top of the valence band in the bulk material (see fig. 2), and hence we have

$$I_p \simeq \frac{4 r_b \mu_p (2\boldsymbol{m})^{3/2} (\pi \boldsymbol{k} T)^{5/2}}{\boldsymbol{h}^3} \exp\left(-\frac{\phi_p}{\boldsymbol{k} T}\right) \exp\left[\left(\frac{\boldsymbol{e} V_e}{\boldsymbol{k} T}\right) - 1\right] \qquad \ldots\ldots(17)$$

(*b*) $p_b \gg n_0$ *and so* $p_a \gg n_0$.

In this case eqn. (14) reduces to

$$V_e \simeq \frac{\boldsymbol{k} T}{\boldsymbol{e}} \log\left(\frac{p_a{}^2}{p_0 n_0}\right).$$

Substituting from (9), expressing n_0 and p_0 in terms of ϕ_n and ϕ_p (see fig. 2) by eqn. (16) and a similar equation for electrons, and finally replacing $\phi_n + \phi_p$ by E_0, the energy width of the forbidden gap, we have

$$I_p \simeq \frac{8 r_b \mu_p (2\boldsymbol{m})^{3/2} (\pi \boldsymbol{k} T)^{5/2}}{\boldsymbol{h}^3} \exp\left(-\frac{E_0}{2\boldsymbol{k} T}\right) \exp\left(\frac{\boldsymbol{e} V_e}{2\boldsymbol{k} T}\right). \qquad \ldots\ldots(18)$$

Equation (18) has been obtained in the condition $p_a \gg n_0$, which corresponds to the case when a large forward current is flowing. When p_a becomes sufficiently large, however, a condition is reached when the assumption of unit injection ratio becomes untenable. The neutrality condition requires that the increase in hole concentration at the barrier p_a shall be accompanied by an equivalent increase in electron concentration, and the electron current over the barrier will eventually become appreciable. The condition for this to occur has been discussed by Bardeen and Brattain (1949). The expression $I_n/I_p = I_p/I_{crit}$, where

$$I_{crit} = \frac{64\pi^3 \boldsymbol{m}^2 \mu^2 \boldsymbol{k}^3 T^3}{\boldsymbol{e} \boldsymbol{h}^3} \exp\left[-\frac{E_0 - \phi}{\boldsymbol{k} T}\right] \text{amp},$$

relates the magnitude of the electron and hole currents I_n and I_p when $I_n \ll I_p$; it serves to indicate that the analysis will break down in germanium, for example, unless the condition

$$I_p \ll 0{\cdot}07 \exp[-(E_0 - \phi)/\boldsymbol{k} T] \text{ amp} \qquad \ldots\ldots(19)$$

is satisfied.

To summarize, then, it appears that, rather than make a distinction between barrier and spreading resistance, we may conveniently regard the system as a whole. If this is done, a graph of log I plotted against V_e may be expected to show the following behaviour:

(*a*) If the condition $p_0 \ll p_a \ll n_0$ can be satisfied, then for values of I satisfying it the curve should have slope $\boldsymbol{e}/\boldsymbol{k}T$, from eqn. (17).

(*b*) As I increases through the condition $p_a = n_0$ to $p_a \gg n_0$, the slope should fall to $\boldsymbol{e}/2\boldsymbol{k}T$.

(*c*) The analysis should break down for large currents, as the electron flow becomes appreciable.

An indication of the manner in which the voltage–current relationship is likely to be altered if the effects of bulk recombination are considered may be obtained as follows. Unless the lifetime of injected carriers is extremely short, it can reasonably be assumed that recombination within the barrier may be neglected, and so for a given current the hole concentration at the edge of the barrier p_b is unchanged. Hence n_b is also unchanged. The decay of holes at larger values of r now requires a flow of electrons radially inwards to replace those lost by recombination. The potential across the bulk semiconductor must therefore be greater than that given in eqn. (3), since this is the voltage required to maintain an electron drift current only just equal and opposite to the electron diffusion current. Thus if i_n is not everywhere zero, V_e must be greater than the corresponding value in eqns. (17) or (18) in order to maintain the same current, and so the current will increase less rapidly with applied voltage.

§ 3. COMPARISON OF THEORY WITH EXPERIMENT

3.1. *The Forward Voltage–Current Characteristic*

Figure 3 shows current–voltage characteristics for forward voltages, plotted semi-logarithmically, for four germanium point contact rectifiers, taken with a tungsten contact, 'formed' by the passage of 100 mA in the forward direction.

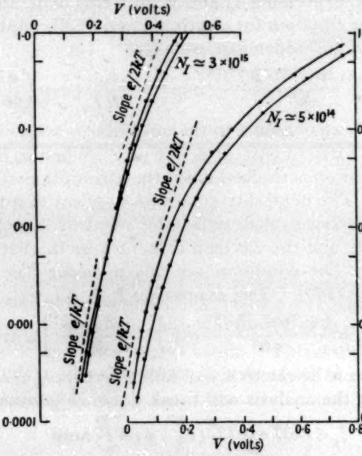


Fig. 3. Forward characteristics of n-type Ge point contact rectifiers, for materials of two different impurity concentrations.

The points were made by etching tungsten wire of diameter two-thousandths of an inch, and the contact radius was estimated as being one-tenth of the radius of the wire, giving $r_b \simeq 3\times10^{-4}$ cm. The curves are for materials of two different impurity contents, estimated from Hall effect and conductivity measurements; using the appropriate values of n_0 and r_b the values of I giving $p_a = n_0$ could be estimated from (7) and (8) for each curve. In the diagram lines of slope $\mathbf{e}/\boldsymbol{k}T$ and $\mathbf{e}/2\boldsymbol{k}T$ are shown in each case in the range of currents in which the appropriate conditions given in the previous section are expected to hold. Conditions (8), (9) and (19) clearly cannot be satisfied over a wide range of currents, and so straight lines with a small transition region could in any case not be expected.

In each case the slope of the experimental curves has fallen below $\mathbf{e}/2\mathbf{k}T$ by the time a current of 1 mA is reached. This might be interpreted as a fall of injection ratio even at these relatively small currents. That condition (19) should break down for currents not greater than of order 1 mA can be explained quantitatively on the following basis. An injection ratio approaching unity is expected to occur when the barrier is of such height that the hole density at the metal–semiconductor contact is large compared with the free electron concentration in the interior. For a semiconductor in the exhaustion range this implies that the hole concentration at the contact is large compared with the donor concentration and hence that the space charge of holes in the barrier near the metal is dominant rather than negligible, as is assumed in the Schottky model. This gives rise to a narrowing and steepening of the barrier, with a consequent increase in the electron current due to image force and tunnel effect. The whole effect may be regarded as equivalent to a reduction in the effective value of the barrier height ϕ in eqn. (19). It is hoped to discuss this effect in more detail in a further publication.

3.2. *Measurements on Contacts under Illumination*

The above analysis has indicated that on the normal model adopted a variable slope, falling below $\mathbf{e}/\mathbf{k}T$ as the current increases, is to be expected when log I is plotted against the total external applied voltage V_e, although log I is still related to the voltage across the barrier, $V_e - V_b$, by a slope of $\mathbf{e}/\mathbf{k}T$, as indicated by eqn. (13).

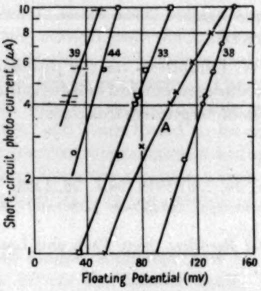


Fig. 4. Measurement of floating potential and short-circuit photo-current for metal–n-type Ge point contact under varying illumination. (Slopes are indicated in volt^{-1}.)

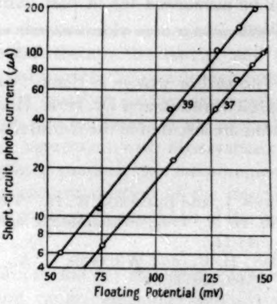


Fig. 5. Measurement of floating potential and short-circuit photo-current for varying distance between illuminated region and point contact. (Fixed point, movable shadow edge.) (Slopes are indicated in volt^{-1}.)

A direct experimental check of the validity of eqn. (13) has been attempted as follows. The hole concentration near the point contact was varied by subjecting the surface to various conditions of illumination described below. In these circumstances the short-circuit photocurrent (zero applied voltage) is expected to be directly proportional to the added hole concentration near the barrier p_a.

Further, neglecting the space charges in the bulk material due to the imperfect balancing of electron and hole diffusion currents, the measured open-circuit photovoltage V_L should be just that voltage *across the barrier* required to produce a forward current equal and opposite to the previously measured photocurrent I_L. Therefore a plot of log I_L against V_L in the region $p_a \gg p_0$ should provide a test of the exponential factor in eqn. (13).

Two types of measurements have been analysed in this way. Figure 4 shows a number of sets of results obtained by flooding the surface round a point contact with light from a tungsten lamp, the intensity being varied by the insertion of neutral filters. The corresponding slopes are indicated in the diagram. In the examination of a number of such sets of readings an occasional one of the type A, showing marked curvature, has been obtained, but nothing approximating to a straight line of slope less than 30 volt^{-1}.

In the second type of test the surface of the germanium was divided into light and dark regions by a sharp shadow edge, the distance of which from the point contact could be varied. I_L and V_L were again measured, and are plotted semilogarithmically in fig. 5. The slopes are in good agreement with the theoretical value of 40 volt^{-1}.

§4. CONCLUSION

An expression has been obtained, with stated simplifying assumptions, for the current–voltage characteristic of a metal–semiconductor point contact rectifier when injection occurs. Agreement with experimental results is satisfactory in view of the approximations made. It appears in particular that the discrepancy in the logarithmic slope encountered in the one carrier model may disappear, or at least be accounted for in part, when injection effects occur and are taken into account.

ACKNOWLEDGMENTS

The author wishes to thank Professor R. W. Ditchburn for the provision of research facilities, and Dr. H. K. Henisch for encouragement and helpful criticism. Thanks are also due to the Admiralty for permission to publish this paper.

REFERENCES

BARDEEN, J., and BRATTAIN, W. H., 1948, *Phys. Rev.*, **74,** 230; 1949, *Ibid.*, **75,** 1208.

BETHE, H. A., 1942, *Massachusetts Institute of Technology, Radiation Laboratory Report* 43/12.

TORREY, H. C., and WHITMER, C. A., 1948, *Crystal Rectifiers* (New York and London : McGraw-Hill).

Forward Characteristic of Injecting Area Contacts on Germanium

By H. K. HENISCH and F. D. MORTEN
Physics Research Laboratory, University of Reading

Communicated by R. W. Ditchburn; *MS. received 29th May* 1953

Abstract. A method is described whereby the voltage–current relation of the barrier associated with an injecting area contact on germanium can be investigated. The method involves capacitance measurements in the presence of injected minority carriers. The effect of the series resistance is eliminated. The results show good agreement with theory.

---

§ 1. INTRODUCTION

In a recent paper, Banbury (1953) has shown that the forward characteristic of an injecting point contact rectifier can be represented by the following expressions which are approximations for two limiting conditions of current flow

$$I \simeq I_0[\exp \boldsymbol{e}U/\boldsymbol{k}T - 1] \qquad \ldots\ldots(1)$$

(concentration of injected carriers small),

$$I \simeq I_0' \exp(\boldsymbol{e}U/2\boldsymbol{k}T) \qquad \ldots\ldots(2)$$

(concentration of injected carriers large),
in which U is the total voltage applied across the assembly as a whole, i.e. the rectifying contact and the bulk material. In this theory, the barrier and the spreading resistance which is modulated by injected minority carriers are considered as part of the same system. The expressions are based on certain assumptions, i.e. unit injection ratio, negligible recombination and fulfilment of the neutrality condition, and were found to be in qualitative agreement with observations of voltage–current characteristics and with the results of photoelectric measurements. It follows from the treatment that the voltage–current relation of the injecting contact itself (i.e. in the absence of a spreading resistance), should be of the form

$$I = I_0[\exp \boldsymbol{e}V/\boldsymbol{k}T - 1] \qquad \ldots\ldots(3)$$

for both ranges of current flow in which eqns. (1) and (2) are valid. V is the voltage across the barrier itself. In the first range, eqns. (1) and (3) are identical ($V \simeq U$) within the limits of the approximation, in the second range $V < U$. This is the type of expression originally suggested by the diode theory, but it should be noted that the diode theory is not involved in the analysis of the injected current.

The voltage U can be directly measured but the voltage V which appears across the barrier can only be indirectly determined. The conventional method of correcting for a linear series resistance is not applicable to injecting contacts and, indeed, the concept of specific resistivity becomes meaningless in the presence of pronounced diffusion effects. In the following section, a method is described whereby the effect of the resistance in series with the barrier due to the bulk material can be eliminated and relation (3) verified.

§2. EXPERIMENTAL PROCEDURE

The present measurements refer to n-type germanium of 4·2 ohm cm resistivity and p-type germanium of 12 ohm cm resistivity. Rectifying electrodes of gold or silver of diameter 0·1 cm were applied by evaporation *in vacuo*, and low resistance base electrodes were soldered.

The principle of the method is illustrated in fig. 1. The a.c. impedance is measured by the bridge circuit (kindly placed at our disposal by Messrs. Standard Telecommunication Laboratories Ltd. and to be described elsewhere). The capacitive component is effectively that of the barrier layer at the test electrode. Illumination of the germanium slab by white light produces hole and electron pairs which diffuse towards the test electrode and, under open circuit conditions, establish a floating potential. This can be measured by means of a potentiometer and the contact capacitance can be determined under open circuit conditions (zero direct current). The injected minority carriers bias the contact in the forward direction, but since no current is flowing, no potential drop is incurred across the resistance of the germanium in series with the barrier. There will be a diffusion voltage across the thickness of the slab, but this is expected to be small except for very intense illumination. The voltage developed across the soldered base electrode (normally shielded from direct illumination) never exceeded 1 mv. It is thus possible to obtain the experimental relation between contact capacitance and the floating potential, which is taken to be the voltage actually across the barrier. Similarly, in the conventional way, it is possible to measure the contact capacitance as a function of total forward applied voltage and current in the absence of illumination. If the capacitance of the barrier is taken as a criterion of the voltage which appears across it, a comparison between these two relations yields the voltages which are across the barrier and the series resistance respectively for any given current. The capacitance was measured at a frequency of 100 kc/s and a check was made to confirm that its value was not markedly frequency dependent. The a.c. signal across the rectifier was kept low enough for the balance to be independent of its magnitude.

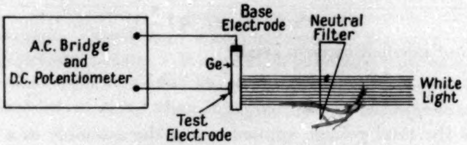


Fig. 1. Schematic diagram of apparatus.

§3. RESULTS

Figure 2 gives typical relations between the intensity of illumination (in arbitrary units) and the corresponding floating potential. If the added hole concentration is assumed proportional to light intensity L, the relation $\ln(1+BL)=\boldsymbol{e}V_{\mathrm{f}}/\boldsymbol{k}T$, where B is constant for a given contact, should be obeyed (Banbury 1952). For sufficiently high values of BL, V_{f} should thus be a linear function of $\ln L$ of slope $\boldsymbol{e}/\boldsymbol{k}T$. Such a linear relation is obtained for the n-type samples, with an average slope of $37 \pm 1{\cdot}5\,\mathrm{v}^{-1}$. For the p-type samples, which gave poor rectifiers and

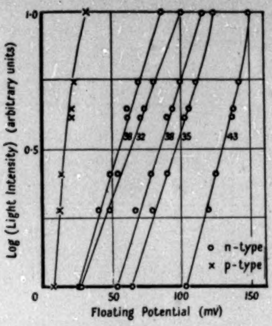


Fig. 2. Measurement of floating potential under varying illumination (slopes marked in volt^{-1}).

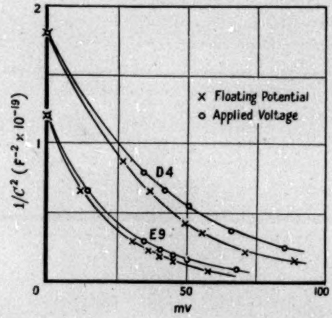


Fig. 3. Typical relations between (capacity)$^{-2}$ and floating potential and applied voltage.

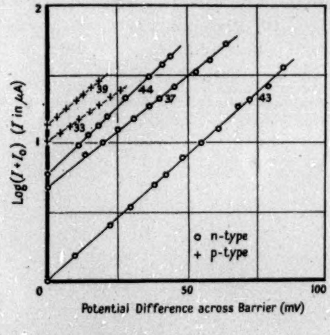


Fig. 4. Current–voltage relationship of the barrier for n-type and p-type rectifiers (slopes marked in volt^{-1}).

developed low photovoltages, the simple relation is not obeyed for the same light intensities owing to a low value of B.

Figure 3 shows the behaviour of the contact capacitance for various floating potentials and externally applied voltages. Since the current for each externally applied voltage is known, these results lead to the voltage–current characteristics of the barrier itself. Figure 4 represents this in the form of a semi-logarithmic plot; here $I = I_0[\exp(\boldsymbol{e}V_f/\boldsymbol{k}T) - 1]$. The values of I_0 are chosen so as to give the best straight line. It will be seen that the slopes, which have a mean value of $39 \pm 2\,\mathrm{v}^{-1}$, are in good agreement with the value of $\boldsymbol{e}/\boldsymbol{k}T$ at room temperature, namely $40\,\mathrm{v}^{-1}$. This is true for the n-type and p-type specimens and is in accordance with eqn. (3). The relation between the magnitude of the series resistance and the current flowing can also be evaluated, but is of no special interest in view of the complicated geometry of the system.

§4. DISCUSSION

For simplicity, Banbury's analysis of the rectifier characteristics and the effect of the spreading resistance involved the assumption that the injection ratio is unity. This assumption may not be satisfied for the present area contacts, but this does not affect the interpretation of the results in fig. 4. The effect of the series resistance has been eliminated and as far as the barrier alone is concerned any part of the forward current which may be carried by *majority* carriers is expected to obey an equation of the same form as (3), on the basis of the normal diode theory. There is thus good agreement between theory and experiment as far as the voltage current characteristic of the contact itself is concerned. This is shown by the slopes in fig. 4, which are in considerably better agreement with the room temperature value of $\boldsymbol{e}/\boldsymbol{k}T$ than those previously recorded in the literature. This improved agreement is also consistent with the relation between floating potential and light intensity, as shown in fig. 2. Various hypotheses could be invoked to account for the remaining discrepancies, but this does not appear worth while at the present stage.

ACKNOWLEDGMENTS

Grateful acknowledgment is made to Professor R. W. Ditchburn for placing facilities at our disposal, to Dr. P. C. Banbury for valuable discussion, to the Royal Naval Scientific Service for their support, and to the Admiralty for permission to publish this paper.

REFERENCES

BANBURY, P. C., 1952, *Thesis*, University of Reading; 1953, *Proc. Phys. Soc.* B, **66**, 833.

A Study of Carrier Injecting Properties of Emitter Contacts and Light Spots at Normal and Moderately Elevated Temperatures

By C. A. HOGARTH
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 1st June 1953.

Abstract. The method of determining the injection efficiency γ of an emitter contact (described previously by Shockley, Pearson and Haynes) is discussed and some of the difficulties of the method and the interpretation of experimental results are described. The effects of a fine spot of white light when used as an emitter and hence as a conductivity modulator are investigated and an equivalent minority carrier current for a given optical assembly can be determined. This procedure suggests an experimental method for the determination of γ as a function of temperature. A simple theory for γ in terms of forbidden energy gap, depth of Fermi level, and barrier height ϕ is given and applied to the determination of barrier heights at room temperature, at elevated temperatures, and when an emitter contact on germanium is illuminated. Theoretical curves relating γ and ϕ are presented for Ge and Si of various impurity concentrations. The experiments carried out at elevated temperatures suggest that the surface states on germanium lie at the top of the filled band and that their density is of order 10^{10}–10^{11} cm^{-2}.

---

§ 1. INTRODUCTION

TRANSISTOR materials (Ge, Si, PbS, and so on) have the unique property that under suitable conditions of excitation a greater number of free charge carriers can exist in the material, at a given temperature and pressure, than in the unexcited material at the same temperature and pressure (for a general reference see Shockley 1950). One method of carrier injection is to pass forward current through a whisker contact on the material; another method consists of focusing a beam of light of suitable wavelength on to the crystal. Using either method, carriers of both signs (electrons and holes) are produced in pairs in the material, but it is the minority carriers, i.e. those of sign opposite to those normally present in the material (as a result of ionization of impurities or of structural defects), which concern us in transistor action. The electrons and holes can be separated at a suitable potential barrier region and can be accelerated in opposite directions by an electric field. There is, however, a strong tendency for the recombination of holes and electrons, and when the source of excitation is removed the number of additional carriers decreases exponentially with time. The carrier lifetime τ which may be determined from this decrease is affected by trapping centres such as defects in the crystal and impurities, and in the case of germanium τ may be as high as a millisecond or as low as a tenth of a microsecond. In general the value of τ at room temperature is much lower at a crystal surface than in the interior.

Shockley, Pearson and Haynes (1949) have defined the efficiency γ of an emitter point contact as the fraction of the emitter forward current carried by minority carriers. Measurements of this quantity are described for n-type single crystal germanium and enable one to assess the expected values of γ after a given surface treatment. Furthermore, if the conductance of a germanium filament is modulated by means of a fine spot of white light effects very similar to those obtained with an emitter point are observed. In their studies of the drift mobility of holes in n-type germanium Lawrance and Gibson (1952) showed that by using a pulsed sweeping field the same result was obtained whether a d.c. emitter or a steady white light spot was used. Measurements made with light spots have enabled an 'equivalent hole current' for a light spot to be obtained, and thus a calibrated source of injected carriers may be obtained and used at various temperatures as a means of determining γ at those temperatures by comparison with the behaviour of an emitter contact. This avoids the difficulties associated with mobility changes, trapping effects, and so on, which will be identical for carriers injected either optically or electrically.

It was believed that the emitter point efficiency at a given temperature and pressure should be determined solely by the barrier height ϕ, forbidden energy gap E, and Fermi energy ϵ. The mathematical formulation of this was carried out with considerable assistance from Dr. L. Pincherle, and the resultant equation was found to give a simple relation between γ, ϕ, E and ϵ. From the measured value of γ on a particular material it is then possible to calculate ϕ, which is of great importance in any study of transistor materials. It is believed that this method of determining ϕ is more reliable than any other method previously described for a surface state Schottky barrier.

The value of γ has been determined for a point contact on germanium in the absence and in the presence of a fine spot of white light around the point. The decrease in γ observed on illumination may be interpreted as a decrease in ϕ, and thus one feature of the barrier theory of photoconductivity is confirmed directly. The value of γ at moderately elevated temperatures is measured and the derived variation of ϕ as a function of temperature is calculated. This enables the depth and concentration of surface states to be estimated.

The experimental results enable some general remarks to be made concerning the operation and manufacture of transistors.

§ 2. THEORY OF EFFICIENCY OF EMITTER IN TERMS OF BARRIER HEIGHT, WIDTH OF FORBIDDEN ENERGY BAND, AND DEPTH OF FERMI LEVEL

The following mathematical formulation was largely the work of Dr. L. Pincherle.*

Consider an n-type semiconductor–metal contact as shown in fig. 1. ϵ is the depth of the Fermi level from the conduction band, ϕ the height of the Schottky barrier, E the forbidden energy band, and V the applied potential (assumed small) which causes forward current to flow. We make the assumption that the effective masses of electrons and holes are the same. This is reasonable since the electron and hole mobilities are of the same order of magnitude. Strictly the analysis should be extended by the introduction of quasi-Fermi levels as used by Shockley (1949) in his theory of p–n junctions, since the semiconductor is in a state of excitation. It is believed, however, that if the applied voltage V is small and all electron and hole transitions are considered to take place from the *interior* of the metal to the *interior* of the semiconductor and vice versa, then the analysis based on the equilibrium diagram given in fig. 1 is applicable to this case. We may regard it as a zero-current approximation.

---

* Subsequent to the preparation of this paper it was brought to the author's notice that a similar equation had been deduced independently by Dr. P. C. Banbury.

Let n_0 and p_0 be the concentrations of electrons and holes respectively at the Fermi level. Then there will be a flow of electrons and of holes from the interior of the semiconductor to the interior of the metal and vice versa. The current density due to electrons flowing from the interior of the semiconductor to the interior of the metal is $Kn_0 \exp\{-(\phi+\epsilon-V)/\boldsymbol{k}T\}$. The current density due to electrons flowing from the interior of the metal to the interior of the semiconductor is $Kn_0 \exp\{-(\phi+\epsilon)/\boldsymbol{k}T\}$. The current density due to holes flowing from the interior of the metal into the interior of the semiconductor is $Kp_0 \exp\{-(E-\epsilon-V)/\boldsymbol{k}T\}$. The current density due to holes flowing from the interior of the semiconductor into the interior of the metal is $Kp_0 \exp\{-(E-\epsilon)/\boldsymbol{k}T\}$. We may define γ as

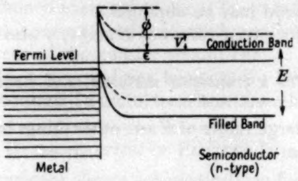


Fig. 1. Energy-level diagram for metal–semiconductor contact.

$$\gamma=\frac{\text{forward current carried by holes}}{\text{total forward current}},$$

and furthermore we have that at the Fermi level $n_0=p_0$. Thus

$$\gamma=\frac{\exp\{-(E-\epsilon-V)/\boldsymbol{k}T\}-\exp\{-(E-\epsilon)\boldsymbol{k}T\}}{\exp\{-(E-\epsilon-V)/\boldsymbol{k}T\}-\exp\{-(E-\epsilon)/\boldsymbol{k}T\}+\exp\{-(\phi+\epsilon-V)/\boldsymbol{k}T\}-\exp\{-(\phi+\epsilon)/\boldsymbol{k}T\}}.$$

Hence, on simplification,

$$\gamma=\frac{1}{1+\exp\{(E-\phi-2\epsilon)/\boldsymbol{k}T\}}. \qquad \ldots\ldots(1)$$

The equation may be readily transposed so as to give ϕ in terms of γ, the result being

$$\phi=E-2\epsilon-\boldsymbol{k}T\ln\{(1-\gamma)/\gamma\}. \qquad \ldots\ldots(2)$$

Thus if γ is measured for a contact on a transistor material for which E, N_0 the impurity concentration and $\Delta\epsilon$ the impurity activation energy are known, ϵ may be calculated and ϕ determined uniquely. This method is considered to be much more reliable than methods previously suggested (see for example Henisch 1949, Torrey and Whitmer 1948), i.e. (i) extrapolation of the linear part of the forward rectification characteristic to the voltage axis, (ii) use of the variation of capacitance with applied voltage, (iii) use of the variation of reverse saturation current i_s with temperature T and of the equation

$$i=i_s \exp\{-e(\phi+\epsilon)/\boldsymbol{k}T\}. \qquad \ldots\ldots(3)$$

Methods (i) and (ii) fail because for point contacts on transistor materials the diffusion theory on which these methods are based no longer applies. Method (iii) fails not only because carrier injection modifies the diode theory equations on which eqn. (3) is based (as described above), but more seriously because ϕ itself is expected to vary with temperature as described in § 7. Thus ϕ cannot be determined from a variation with temperature of some other property of the contact. The determination of ϕ from measurements of γ suffers from none of the above disadvantages since it depends upon the essential feature of carrier injection and, as shown in § 7, the determination of γ as a function of temperature enables the corresponding variation of ϕ to be accurately determined.

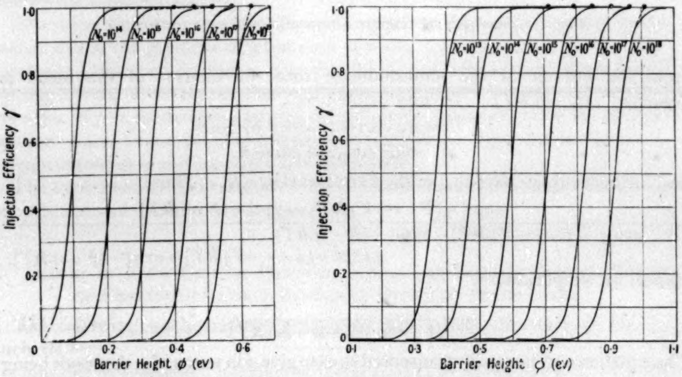
It is instructive to examine eqns. (1) and (2) to test that they agree with experimental facts concerning carrier injection.

(i) It is obvious that eqn. (1), which is of the same form as the well-known Fermi function, cannot be solved for $\gamma=0$ or $\gamma=1$ although solutions as near to these limits as are required may be obtained.

(ii) As ϕ increases for given values of E and ϵ, γ also increases, as expected physically.

(iii) As the purity of a transistor material, and hence ϵ, increases, γ also increases, as expected.

(iv) Materials with large values of E are, other things being equal, less favourable for carrier injection.



(*a*) (*b*)

Fig. 2. Injection efficiency as a function of barrier height. (*a*) Germanium (impurities fully ionized). (*b*) Silicon (impurity activation energy 0·08 ev). $T=290°$K.

The general form of eqn. (1) is illustrated in figs. 2(*a*) and (*b*) for germanium (n-type), $E=0{\cdot}72$ ev, and silicon (p-type), $E=1{\cdot}1$ ev. To determine ϕ for $\gamma=0$ and $\gamma=1$ the values are calculated as follows. If $\phi=0$ then there is no potential barrier and presumably no injection so that $\gamma=0$. For the maximum value of $\gamma(=1)$ we should expect the filled band of an n-type material to be raised to the Fermi level at the surface. Considerations of a similar nature would apply to the depression of the conduction band of a p-type semiconductor. Any further lowering of the conduction band or raising of the filled band would require an impossibly high concentration of surface states. In this case then $\phi=\phi_{\max}$, where $\phi_{\max}$ is given by $E-\epsilon$. In both diagrams the curves illustrate the relation between ϕ and γ in the expected manner, but in extreme cases require correction. When N_0 is very small, e.g. $10^{13}\,\mathrm{cm}^{-3}$, there is an error in the calculated value of ϕ at low injection efficiency because the motion of holes in the filled band (for the n-type case) has not been considered in the derivation of ϵ from N_0, Furthermore, when $N_0=10^{18}\,\mathrm{cm}^{-3}$ or greater, ϵ should have been determined from a true Fermi–Dirac function and not from a classical approximation. In materials of the purity used in practical transistors the approximations made are entirely justified.

It is further of interest to examine the condition for transistor action given by Bardeen and Brattain (1949), namely $\phi+\epsilon \geqslant \frac{1}{2}E$, for which the limiting condition is $\phi+\epsilon=\frac{1}{2}E$. From an inspection of figs. 2 (*a*) and (*b*) it may be observed that on this basis transistor action is expected to occur for very low values of ϕ. From practical considerations it seems that for a useful transistor assembly at least half the total forward current of the emitter should consist of minority carriers, in which case $\gamma=0{\cdot}5$. Thus the condition becomes $\phi+2\epsilon \geqslant E$ which suggests that higher values of ϕ are required if a useful transistor assembly is to be achieved.

§ 3. THEORY OF DETERMINATION OF EMITTER INJECTION EFFICIENCY

Consider the arrangement shown schematically in fig. 3, where the transistor material has a filamentary structure and a one-dimensional analysis may be applied with small resultant error. If the emitter contact is biased forward there will be an increase in conductance as a result of carrier injection, and this may be observed along the length of the filament by means of suitable probes since the injected carriers travel down the filament under the influence of the sweeping current I_b. In the absence of recombination of electrons and holes in their passage along the filament it has been shown by Shockley, Pearson and Haynes (1949) that for an n-type semiconductor γ is given by the equation

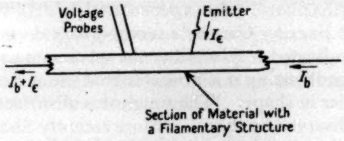


Fig. 3. Schematic arrangement.

$$\gamma=\frac{(1-G_0/G)(I_b+I_\epsilon)}{(1+b)I_\epsilon} \qquad \ldots\ldots(4)$$

where G_0 is the conductance of the filament between the probes in the absence of injection ($I_\epsilon=0$), G is the conductance of the filament in the presence of injection, I_b the sweeping current, I_ϵ the emitter current, b the mobility ratio μ_e/μ_h; μ_e, μ_h are the drift mobilities of an electron and a hole. (For p-type semiconductors b is replaced by $1/b$ in eqn. (4).)

Since the effect of conductivity modulation at a distance x from the emitter will be reduced as a result of recombination, eqn. (4) will only give γ_e, i.e. an effective efficiency. To correct for this, readings are taken for different values of I_b, and the sweeping field along the filament is also measured. Then if V is the potential fall along the distance x and μ is the mobility of a minority carrier, the transit time t is $x^2/\mu V$. Then by plotting log γ_e against t and extrapolating to $t=0$ we obtain a true value for γ for the case of zero recombination. The slope of the line produced in this way gives the reciprocal of the minority carrier lifetime τ and thus provides a further method for the determination of γ. If a rapid determination of γ is all that is required it is sufficient to plot log γ_e against $1/V$ and extrapolate to $1/V=0$.

The theory as developed by Shockley *et al.* (1949) involves an important approximation which is not always justified in practical cases. In the derivation, the conductance in the presence of injection is taken to be a linear function of the carrier concentrations and the sweeping field, and the terms involving diffusion of carriers are neglected. For the lower values of sweeping currents the diffusion term may in fact be quite large or even predominate. This imposes practical limitations on the experimental arrangements.

From the form of eqn. (4) it may be observed that the accuracy of γ_e is very much affected by the quantity $1-G_0/G$. In practice it has been found that G_0/G is often of order 0·9, so that $1-G_0/G$ is about 0·1. Thus a small error (say 1%) in G_0/G may cause a much larger error (say 10%) in $(1-G_0/G)$. For this reason G_0 and G must be determined with high accuracy. Even so the errors in γ_e, and hence in γ, are likely to be of order 5–10%.

There are very few actual measurements of γ reported in the literature. In the measuremens of Shockley *et al.* (1949) the value of the factor $1+b$ was taken as 2·5 instead of the currently accepted value of 3·1. This would have the effect of reducing the measured values of γ from 1 to 0·8 for n-type Ge and from 0·6 to 0·48 for p-type Ge. Furthermore, if the experimental points given by the above authors are fitted to straight lines by Cauchy's method they do not pass exactly through $\gamma=1$ and $\gamma=0{\cdot}6$ as indicated. Also, since the specimens used by the above authors were prepared by sandblasting it is probable that the surfaces were rough and not accurately rectangular in shape. The roughness of surface would probably cause the low values of τ observed by them. More recently Shockley (1950, p. 346) has stated that it is possible to obtain a range of values of γ for n-type Ge, but details of the methods and results have not been published.

§ 4. EXPERIMENTAL ARRANGEMENTS

As presented above, the determination of γ could be performed by a.c., d.c. or pulsed signals. Although by means of applied pulses higher fields may be achieved and hence smaller transit times occur, it was considered that the worst source of error would arise from the voltage and current measurements to obtain G and G_0. The experiments were therefore carried out using d.c., since by this method the voltages could be measured with a standard potentiometer of high accuracy and the currents could be readily checked with the aid of high quality d.c. meters. The total current through the specimen was maintained sufficiently high to make the diffusion effects negligible and sufficiently low to avoid effects due to Joule heating.

4.1. *Experimental Specimens*

Experiments were carried out on various specimens of n-type single crystal germanium having resistivities of 2·2, 7 and 20 ohm cm and carrier lifetimes from

5 to 50 microseconds, the latter quantity being very dependent on the surface treatment given the specimens. The specimens were cut roughly to size by means of a special tungsten-wire–carborundum saw, were ground with fine carborundum and water, and were dried, polished, suitably etched and washed. The final dimensions were of order 0·1 to 0·4 mm × 0·1 to 0·4 mm × 1 to 2 cm. The total resistances of the filaments were thus of order 1000–20 000 ohms. They were of good geometrical shape, clean, and had a bright appearance. They were mounted on Perspex blocks in the manner shown in fig. 4 (*a*).

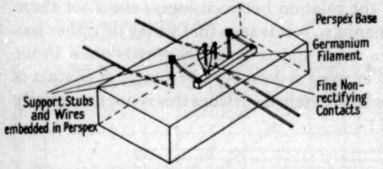

Perspex Base
Germanium Filament
Fine Non-rectifying Contacts
Support Stubs and Wires embedded in Perspex

Fig. 4 (*a*). Perspective view of crystal mounting.

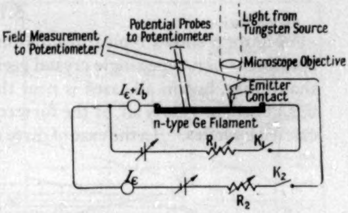

Field Measurement to Potentiometer
Potential Probes to Potentiometer
Light from Tungsten Source
Microscope Objective
Emitter Contact
$I_\varepsilon + I_b$
n-type Ge Filament
R_1
K_1
I_ε
K_2
R_2

Fig. 4 (*b*). Circuit diagram.

4.2. *Soldering Procedure*

The end contacts were of large area and made with a small soldering iron after the application of a small amount of lactic acid as a flux.

The potential probes were soldered to the filament using a micro-soldering technique developed largely by W. H. Mitchell (to be published). By this technique it was found possible to make a soldered junction of area no greater than that covered by the 0·002 inch platinum wire used. Furthermore, junctions made in this way remained non-rectifying down to low temperatures and enabled experiments at various temperatures to be carried out with confidence. By the use of contacts of this kind it is thought that the effects of floating contact potentials described by Bardeen (1950) have been avoided.

4.3. *Measuring Method*

The circuit is shown in fig. 4 (*b*) and was used as follows. K_1 was closed and the sweeping current was adjusted to a suitable value by means of resistance R_1. The potential difference across the probes ΔV_0 was measured and the value of the sweeping current noted. K_2 was then closed and the emitter current I_ε adjusted to a convenient value less than I_b. The new potential difference ΔV, I_ε and $I_\varepsilon + I_b$ were noted. The potential fall along the filament between emitter and probe was measured so as to estimate the sweeping field. Then

$$\gamma_e = \frac{\{1 - [I_b \Delta V/(I_b + I_\varepsilon)\Delta V_0]\}}{(1+b)} \frac{(I_b + I_\varepsilon)}{I_\varepsilon},$$

and by making measurements for various values of I_b, γ could be computed as described above.

4.4. *Emitter Points and Light Spots*

The emitter points were made from 0·004 in. phosphor-bronze or tungsten wires pointed in the usual way. They were applied to the rod by means of a micromanipulator and no efforts were made to electroform them. One experiment was carried out using a stretched 0·004 in. phosphor-bronze wire as a line rather than a point emitter.

For optical injection fine spots of light were produced quite simply from a tungsten lamp and microscope objective. By means of an adjustable lens-holder the spot could be readily moved over small distances without upsetting the rest of the arrangement, and could be readily defocused if necessary.

§ 5. Experimental Results

5.1. *General*

Figure 5 shows typical results for the relation between log γ_e and t for three specimens of n-type single crystal germanium. It is seen that γ may be rather less than unity, but in all cases is near this value and is certainly greater than about 0·8, i.e. all, or nearly all, of the forward current at an n-type germanium contact is carried by holes. In the case of dirty or badly etched surfaces this is not necessarily so, and γ can assume smaller values. With the small values of emitter current used no dependence of γ on I_E was observed. The values of lifetime calculated from the graphs are: specimen A, 17 μsec; specimen B, 25 μsec; specimen C, 14 μsec. These values are of the same order of magnitude as those obtained by the light spot method used by Lawrance and Gibson (private communication), and in fact specimen A is the same as that for which the above authors measured a value of 18 μsec. The values of lifetime obtained from these measurements lie between the values for the body and surface lifetimes.

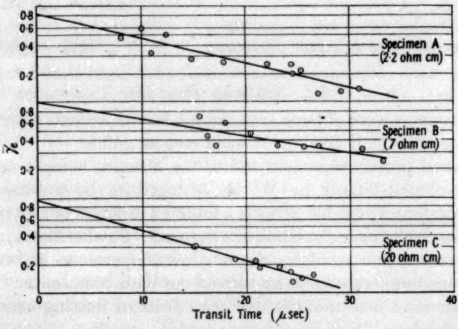


Fig. 5. Determination of emitter point γ for contacts on three specimens of germanium.

The experiment carried out with the stretched wire emitter yielded a value of $\gamma = 0{\cdot}9$, so that this arrangement is very useful in practical devices such as the infra-red modulator devised by Gibson (1953).

5.2. *Calibration of a Light Spot Emitter*

The complete specification of the effect of a light spot emitter would involve a knowledge of the incident radiant energy and other parameters which in practice require fairly complicated measurement. It has been found satisfactory to measure the 'equivalent hole current' for a light spot and then keep the system fixed during the subsequent experiments. The use of this device is described in § 6. Measurements of probe potential difference were made for a given *total* current through the specimen for various values of emitter current and for a light spot emitter. The results for a typical experiment are given in fig. 6. From the curves a plot of emitter current against potential difference can be constructed and the equivalent hole current $I_{h\epsilon}$ calculated. For the filament used in fig. 6 $I_{h\epsilon}$ is 3 mA. For higher resistance filaments this value would be smaller.

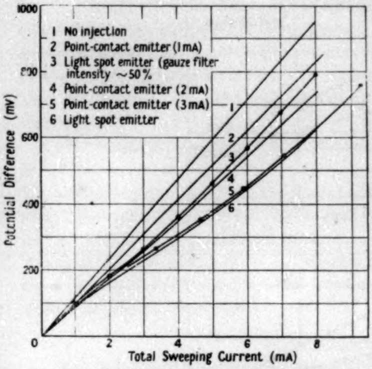


Fig. 6 (*a*). Conductivity modulation by point contact and light spot carrier injection (single crystal 7 ohm cm Ge).

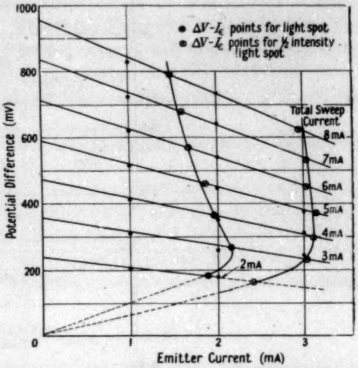


Fig. 6 (*b*). Determination of equivalent hole current for a light spot emitter.

5.3. *Change of γ on Illumination*

Figure 7 shows the effect of illumination on the values of γ for emitter contacts on 2·2 and 7 ohm cm germanium respectively. It is noteworthy that γ decreases markedly upon illumination of the point contact with a light spot of the type described above. There is an apparent increase in τ upon illumination. The theory discussed in § 2 enables us to interpret the change in γ in terms of a change in barrier height $\Delta\phi$. It is to be expected that E will be unaffected by illumination, and that ϵ will also be little affected except possibly within a very small region of the crystal near the surface. In any case the theory for γ concerns transitions from the interior of the crystal to the metal and vice versa. The maximum change will be that in ϕ caused by the excitation of free carriers to empty some of the surface states and hence to decrease γ. If we then assume that γ changes predominantly as a result of the change in ϕ we may estimate $\Delta\phi$ as follows :

(i) 7 ohm cm germanium. γ changed from 0·95 to 0·17, corresponding to a change in ϕ from 0·220 to 0·106 ev, i.e. $\Delta\phi = 0{\cdot}11$ ev.

(ii) 2·2 ohm cm germanium. γ changed from 0·84 to 0·175, corresponding to a change in ϕ of from 0·250 to 0·168 ev, i.e. $\Delta\phi = 0{\cdot}08$ ev.

Thus in each case the decrease in barrier height on illumination is of order 0·1 ev, although the change is rather higher in the case of the specimen of higher resistivity. It is believed that this experiment provides the first direct evidence for the barrier theory of photoconductivity as postulated for example by Gibson (1951, 1952).

5.4. *Change of Carrier Lifetime on Illumination*

The results shown in fig. 7 suggest that τ increases upon illumination of the specimen. This may be explained qualitatively by assuming that a small region around the contact is illuminated and that the relatively shallow traps on the surface of the filament (of the kind suggested by Lawrance (1953)) in this region have been filled by the white light. It is thus less likely for an injected carrier to be trapped and hence τ increases. This interpretation was verified by measuring γ with the light spot defocused slightly and incident on the filament *between* the emitter and the probe contacts. The result is shown in fig. 8, and it may be observed that although γ is relatively unaffected, τ has increased when the surface of the crystal is illuminated, so that presumably more traps have been filled and the probability of surface trapping has been decreased.

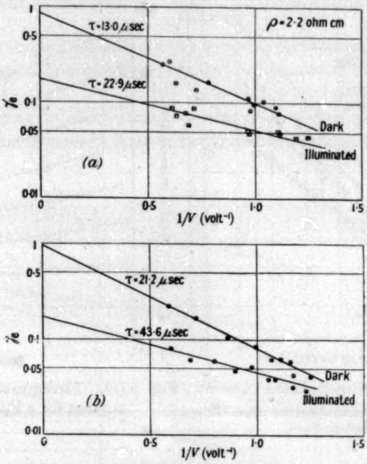


Fig. 7. Variation of γ and τ upon illumination. (*a*) 2·2 ohm cm germanium, (*b*) 7 ohm cm germanium.

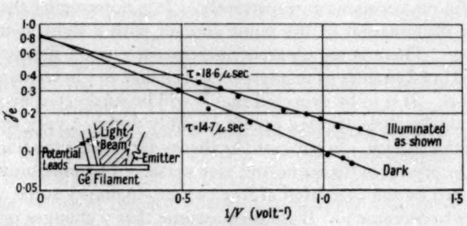


Fig. 8. Variation in lifetime with illumination along the crystal (2·2 ohm cm germanium).

§ 6. EXPERIMENTS AT ELEVATED TEMPERATURES

6.1. *General*

It is known that germanium transistors cease to function efficiently above about 60°C, and it is of interest to assess if possible the relative importance of effects due to the emitter and collector separately since this may affect the design considerations of practical devices. It is known that the efficiency of a collector contact decreases markedly as its reverse resistance decreases, and it is therefore less efficient as a device for collecting holes and acting as a barrier to electrons. Since the saturation current in the reverse direction is an exponential function of temperature a large effect from this cause is to be expected. There is no published information regarding the operation of emitters at elevated temperatures, and hence the following investigation of γ as a function of temperature was made. In addition it was found that the results enabled the location and density of the surface states to be determined.

The ordinary method of determination of γ at elevated temperatures described in § 3 has many difficulties since there will be considerable uncertainty regarding the values of μ_{e}, μ_{h} and b at any temperature. The method used here, in which the effect due to a 'calibrated' light spot is compared with that due to the emitter point, avoids these difficulties since the transit effects for electrons or holes injected either optically or electrically are expected to be identical. (This consideration would be even more important at low temperatures where trapping effects would be expected.)

6.2. *Experimental Method*

The experiment is carried out as follows: (1) the value of γ_0, the efficiency at room temperature, is measured in the usual way; (ii) the equivalent minority carrier current $I_{\mathrm{m}\epsilon}$ of a light spot incident at the emitter contact is measured for various sweeping currents; (iii) the change in conductance between the voltage probes ΔG_{L} for a given value of total sweeping current and the use of the light spot emitter only is determined at some other temperature T; (iv) the change in conductance ΔG_{E} for the same sweeping current, but with the emitter contact passing forward current I_ϵ of similar magnitude to $I_{\mathrm{m}\epsilon}$, is also determined at temperature T. Then it may be simply shown that

$$\Delta G_{\mathrm{L}}/\Delta G_{\mathrm{E}} = I_{\mathrm{m}\epsilon}\gamma_0/I_\epsilon\gamma_T$$

and hence γ_T the emitter point efficiency at temperature T is readily calculated. The method assumes that the quantum efficiency of the material does not vary significantly with temperature and also that the wavelengths of light used are largely well within the lattice absorption band of the material at all the temperatures of measurement. In the case of germanium the first assumption is believed to be correct since the quantum efficiency is not expected to vary more rapidly with temperature than E, the forbidden gap width. The second assumption is also justified since at normal operating temperatures a tungsten filament lamp has its maximum radiation at about $1\cdot 1\mu$, well within the absorption band of germanium.

One advantage of this experimental method is that even if μ_{e}, μ_{h} and b for a given material are not known, so that γ_0 may not be determined, the variation of γ/γ_0 with temperature may still be investigated.

The etched germanium filament was mounted on a small block of Perspex with a very thin phosphor-bronze shim at one end so that any expansion or contraction of the specimen with respect to its support would not introduce strains into the crystal and hence cause incorrect results (Lawrance 1953). Potential probes were soldered on as previously described and a chromel–alumel thermocouple to one end of the filament. The whole was held in a micromanipulator designed by W. H. Mitchell, the emitter contact made to the filament, and its distance from the probes noted. In a typical experiment the spacing of the probes would be about 0·2 to 0·3 mm and the emitter about 2·5 to 3 mm away. The whole assembly was then fitted in a Dewar type cell with a plane window so that a fine light spot could be focused on the filament around the emitter contact. The cell could be evacuated, and heated by means of a resistance element immersed in transformer oil in the centre container.

The electrical measurements were made as described above. Precision in the current–voltage measurements was of great importance since I_ϵ was kept to approximately the same value as $I_{m\epsilon}$ and hence the changes in conductance ΔG_E and ΔG_L would be of the same order of magnitude, possibly differing by 25–50% as γ decreased. Reproducible results could readily be obtained and γ determined as a function of temperature.

6.3. *Results of Experiments*

The effect of elevated temperatures on the value of γ for unformed emitter contacts on n-type germanium is shown in fig. 9. It may be observed that γ remains constant over the temperature range covered with the 2·2 ohm cm specimen, whereas γ for the contact on 7 ohm cm material decreases steadily above about 60°c, at which temperature an appreciable thermal generation of hole-electron pairs would be expected. As the specimen reaches the intrinsic range γ continues to fall. For the 2·2 ohm cm specimen the temperature at which the intrinsic contribution to the conduction becomes important has not been reached. (The point shown at $\gamma = 1{\cdot}06$ for this specimen illustrates the magnitude of the experimental errors in the method.)

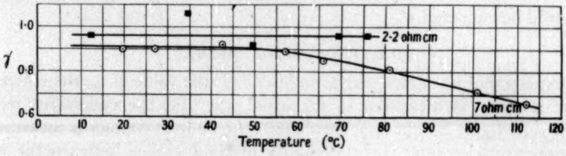


Fig. 9. Temperature variation of emitter point γ for contacts on two specimens of germanium.

§ 7. THEORETICAL INTERPRETATION OF THE TEMPERATURE DEPENDENCE OF γ

The measurements of γ at elevated temperatures enable us to deduce the variation with temperature T of the potential rise at the surface ϕ, by the use of eqn. (2). The appropriate values of E and ϵ were calculated for the particular temperatures concerned. The variation of ϕ with T is shown in fig. 10, and it may be observed that ϕ decreases fairly slowly with temperature.

From values of ϕ at different temperatures we may calculate the position of the surface states with respect to the top of the upper filled band at the surface. The equations used were derived by Bardeen (1947) and have been used by Gibson (1952). The theory employed is developed for an n-type semiconductor but—

mutatis mutandis—applies equally well to p-type material. Let us assume that the surface states occupy a discrete level E_s above the filled band and E_d below the Fermi level. Then

$$E_d = E - \phi - \epsilon - E_s \qquad \ldots\ldots(5)$$

Since the number n of occupied states per unit area of those available n_0 is given by

$$n = n_0[1/\{1+\exp(-E_d/\boldsymbol{k}T)\}], \qquad \ldots\ldots(6)$$

then

$$n = n_0\{1/[1+\exp\{-(E-\phi-\epsilon-E_s)/\boldsymbol{k}T\}]\}. \qquad \ldots\ldots(7)$$

Furthermore, Bardeen (1947) has shown that

$$\phi = 2\pi e n^2/KN_0, \qquad \ldots\ldots(8)$$

where K is the dielectric constant of the material. Thus

$$\left(\phi\frac{KN_0}{2\pi e}\right)^{1/2} = n_0\frac{1}{1+\exp\{-(E-\phi-\epsilon-E_s)/\boldsymbol{k}T)\}},$$

so that

$$\left(\frac{\phi_1}{\phi_2}\right)^{1/2} = \frac{1+\exp\{-(E_2-\phi_2-\epsilon_2-E_s)/\boldsymbol{k}T_2\}}{1+\exp\{-(E_1-\phi_2-\epsilon_1-E_s)/\boldsymbol{k}T_1\}}. \qquad \ldots\ldots(9)$$

Thus from a knowledge of ϕ at two different temperatures E_s can in principle be determined from eqn. (9). An exact solution is not always achievable, but from the values shown in fig. 10 the value of E_s seems to be approximately zero ($\pm 0{\cdot}02$ ev),

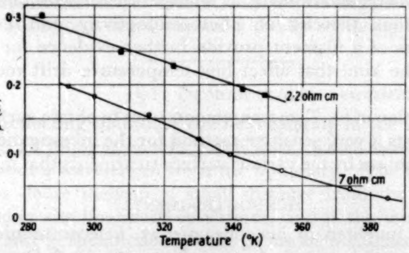


Fig. 10. Temperature variation of barrier height ϕ for two specimens of germanium.

i.e. the surface state level is at the top of the filled band at the surface. For very pure materials a more rigorous calculation (Mrs. A. M. Woodward, private communication) shows that ϕ should be replaced by $\phi + \boldsymbol{k}T\exp\{-(E-\phi-2\epsilon)/\boldsymbol{k}T\}$ in eqns. (8) and (9). The results quoted in this paper were obtained with the use of this correction.

Since at 0°K the value of ϵ is approximately zero, the extrapolated value of ϕ should be $E - E_s$. By extrapolating the linear portion of the graphs in fig. 10 to $T=0$ values of about 0·75 ev are obtained in each case. This confirms that E_s is approximately zero.

The surface state density was calculated from eqns. (7) and (8) with the corrected value of ϕ and the results obtained for n_0 for the 2·2 and 7 ohm cm specimens were $1{\cdot}1 \times 10^{11}$ and $4{\cdot}3 \times 10^{10}$ cm^{-2}. These values are rather lower than that of 10^{12} cm^{-2} suggested by Bardeen (1947). In his calculation, however, the concentration of impurities N_0 was taken as about 10^{17} cm^{-3} instead of the 10^{14}–10^{15} cm^{-3} appropriate to the above specimens. Thus for a given surface rise of potential ϕ the density of surface states would be greater in his case.

§ 8. Discussion and Conclusions

The experimental results suggest that for clean surfaces of n-type single crystal germanium any clean rectifying contact, point or line, will have an injection efficiency of from 0·8 to 1. This implies that in transistor construction there is probably no need to form the emitter electrically so long as clean whisker materials and well-etched crystals are employed. The failure of transistors to operate above 60–70°c may be attributed *in part* to a decrease in emitter point efficiency for devices made from germanium of resistivity 5–6 ohm cm or greater. The chief cause of failure is still expected to be the decrease in reverse resistance of the collector. For germanium of lower resistivity the failure of the collector is likely to be the sole cause of failure of the transistor. (Other effects, such as the rapid decrease in drift mobility at an elevated temperature—about 55°c—reported by Lawrance (1953), will also affect transistor performance.)

The theory outlined provides a useful relation between ϕ and γ and enables ϕ to be determined uniquely once γ is measured for a material of known purity. This method of determination is probably more reliable than any previously described for this class of solid.

It is shown that a light spot can act as a standard source of carrier injection and may be used as a 'sub-standard', thereby enabling γ at temperatures other than room temperature to be determined. The experiments with illuminated contacts substantiate previous theories of photoconductivity, and experiments with illuminated regions of a filament provide further evidence for the existence of surface traps of the kind that affect low temperature drift mobilities and also current gain in transistors.

The interpretation of (γ, T) measurements so as to obtain surface state position and density suggests a very sensitive method for the investigation of the changes produced in germanium by the various surface treatments that may be used.

Acknowledgments

The author is indebted to his colleagues at Telecommunications Research Establishment for much help and advice, and in particular to Drs. L. Pincherle and A. F. Gibson for discussions regarding the theory and to W. H. Mitchell, who mounted the specimens, also to Dr. P. C. Banbury of Reading University for several useful criticisms. He is indebted to the Chief Scientist, Ministry of Supply, and the Controller, H.M. Stationery Office, for permission to publish this paper.

References

Bardeen, J., 1947, *Phys. Rev.*, **71,** 717; 1950, *Bell Syst. Tech. J.*, **29,** 469.

Bardeen, J., and Brattain, W. H., 1949, *Phys. Rev.*, **75,** 1208,

Gibson, A. F., 1951, *Proc. Phys. Soc.* B, **64,** 603; 1952, *Ibid.*, **65,** 196; 1953, *Ibid.*, **66,** 588.

Henisch, H. K., 1949, *Metal Rectifiers* (Oxford : Clarendon Press).

Lawrance, R., 1953, *Phys. Rev.*, **89,** 1295.

Lawrance, R., and Gibson, A. F., 1952, *Proc. Phys. Soc.* B, **65,** 994.

Shockley, W., Pearson, G. L., and Haynes, J. R., 1949, *Bell Syst. Tech. J.*, **28,** 344.

Shockley, W., 1949, *Bell Syst. Tech. J.*, **28,** 435; 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).

Torrey, H. C., and Whitmer, C. A., 1948, *Crystal Rectifiers* (London and New York : McGraw-Hill).

Effect of a Metal Plate on Total Reflection *

BY W. CULSHAW † AND D. S. JONES ‡

† Telecommunications Research Establishment, Ministry of Supply, Great Malvern, Worcs.
‡ Mathematics Department, University of Manchester

MS. received 19*th February* 1953, *and in final form* 8*th June* 1953

Abstract. The phase changes which occur on total reflection are modified when a metal plate is placed near and parallel to a totally reflecting boundary. In particular the phase difference between plane waves of equal amplitude, polarized in and perpendicular to the plane of incidence, can be varied from $-\pi$ through zero and up to the value obtained without the plate. When the refractive index is greater than or equal to $1+\sqrt{2}$, there are two positions of the plate for which the reflected wave is circularly polarized, and in which the electric vector rotates in opposite directions.

Measurements of this effect have been made at a wavelength of 1·25 cm, with a microwave analogue of the optical spectrometer, and a 45° Perspex prism with faces 6 in. square. In view of the limitations due to diffraction the results are in good agreement with the theory, and are an elegant way of showing the presence of the evanescent wave beyond a totally reflecting boundary.

§ 1. INTRODUCTION

At a totally reflecting boundary the field existing in the less dense medium is due to an inhomogeneous plane wave, which is effective at distances up to about a wavelength from the boundary. The presence of a metal plate in the less dense medium a small distance from the boundary may be expected to modify the boundary conditions, and if the incident radiation has components in and perpendicular to the plane of incidence, this modification will be shown by a change in polarization of the reflected waves.

Apart from limitations on aperture size and possible diffraction errors, the investigation of the phenomena is facilitated at wavelengths in the microwave region, where accurate control of the polarization and intensity can also be effected with waveguide techniques. The measurements have thus been made with a microwave analogue of the optical spectrometer, and a 45° totally reflecting Perspex prism, together with an arrangement whereby the plate can be set parallel to the totally reflecting surface and its distance from it accurately determined. The phase change was measured for various distances of the metal plate and the results compared with theory.

§ 2. THEORY

The method of analysis is similar to that given by Stratton (1941). We assume that the transmitting horn provides a plane wave, and neglect any diffraction effects due to the edges of the prism. We also assume that the prism has the permeability of free space with no conductivity, and neglect any reflections which occur at its faces. The metal plate is assumed to be a perfect conductor.

---

* The work described in the paper comprises part of a thesis approved by the University of London for the award of the degree of Ph.D.

A patent application relating to the subject matter of the paper has been filed.

The coordinate system and arrangement of the media are shown in fig. 1, where a plane wave travelling in a medium of refractive index N is incident on a plane interface separating the medium from free space. Parallel to this interface and at a distance d from it is a perfectly conducting plane. The waves existing in the system may be determined from the boundary conditions, viz. that the tangential components of the electric and magnetic intensities are continuous across the planes $y=0$ and $y=d$. With the direction of propagation in the plane $z=0$ this leads to the following results, where the variation with time is omitted.

If the incident wave is polarized with the electric field

$$e_0 \exp\{-ik(x\cos\theta+y\sin\theta)\}$$

perpendicular to the plane of incidence, then the electric field of the reflected wave in the region $y\leqslant 0$, viz. $e_1 \exp\{-ik(x\cos\theta-y\sin\theta)\}$, is given by

$$\frac{e_1}{e_0}=-\frac{(1+\psi)\sin\theta_2-(1-\psi)N\sin\theta}{(1+\psi)\sin\theta_2+(1-\psi)N\sin\theta} \qquad \ldots\ldots(1)$$

where $\cos\theta_2=N\cos\theta$, $\psi=\exp(-2ikd\sin\theta_2)$, $k=2\pi/\lambda$, and λ is the free-space wavelength.

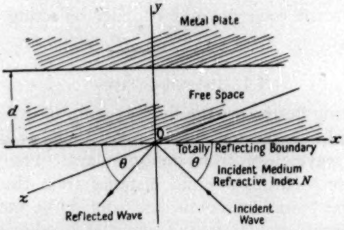


Fig. 1. Coordinate system.

Similarly if the incident wave is polarized with the magnetic field perpendicular to the plane of incidence, the magnetic field of the reflected wave in the region $y\leqslant 0$ is given by

$$\frac{h_1}{h_0}=-\frac{(1-\psi)N\sin\theta_2-(1+\psi)\sin\theta}{(1-\psi)N\sin\theta_2+(1+\psi)\sin\theta} \qquad \ldots\ldots(2)$$

where h_0 is the amplitude of the magnetic field in the incident wave.

Total reflection occurs in the absence of the metal plate when

$$N\cos\theta\geqslant 1 \qquad \ldots\ldots(3)$$

and the proviso that the transmitted wave be exponentially damped requires that

$$\sin\theta_2=-i(N^2\cos^2\theta-1)^{1/2} \qquad \ldots\ldots(4)$$

where the positive sign of the square root is taken. Equations (3) and (4) are not affected by the presence of the metal plate.

With the metal plate present and the incident wave polarized at 45° to the normal to the plane of incidence, let e_0' and e_0'' refer to the electric fields in the incident waves with the electric vector normal and parallel respectively to the plane of incidence. Let $e_0'\exp(i\delta')$ and $e_0''\exp(i\delta'')$ similarly refer to the reflected waves. Then, from eqns. (1), (2) and (4), the phase difference $\delta = \delta' - \delta''$ between the two oscillations is given by

$$\tan \tfrac{1}{2}\delta = \frac{\{N^2(1-\psi)^2 - (1+\psi)^2\}\mu \sin\theta}{N(N^2-1)(1-\psi^2)\cos^2\theta} \qquad \ldots\ldots(5)$$

where $\mu = (N^2 \cos^2\theta - 1)^{1/2}$, and now $\psi = \exp(-2k\mu d)$. As $d \to 0$, $\psi \to 1$, and $\tan\frac{1}{2}\delta \to -\infty$. Also $\tan\frac{1}{2}\delta = 0$ when

$$N^2(1-\psi)^2 = (1+\psi)^2 \qquad \ldots\ldots(6)$$

and so, as d increases, δ increases from $-\pi$, passes through zero and up to the value obtained when the plate is removed.

The reflected wave is circularly polarized when $\tan\frac{1}{2}\delta = \pm 1$, and it follows that the value $\tan\frac{1}{2}\delta = 1$ is not always possible. When $\tan\frac{1}{2}\delta = 1$, eqn. (5) becomes a quadratic in ψ, which when solved gives

$$d = \frac{1}{2k\mu}\log\left\{\frac{N^2 - 1 + \alpha}{N^2 + 1 \pm (4N^2 + \alpha^2)^{1/2}}\right\} \qquad \ldots\ldots(7)$$

where $\alpha = N(N^2-1)\cos^2\theta/(\mu\sin\theta)$.

Since d must be real and positive, it follows that the condition for $\tan\frac{1}{2}\delta = 1$ to occur is

$$N\cos^2\theta \leqslant (\sin\theta)/\mu \qquad \ldots\ldots(8)$$

and that the negative sign must be taken before the radicle. Hence if $2N/(N^2-1) > 1$, eqn. (8) cannot be satisfied for any angle of incidence, and there is no position of the plate at which this type of circular polarization occurs.

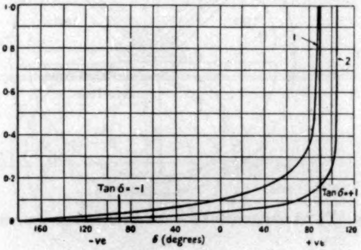


Fig. 2. Phase change δ plotted against distance of plate d for dielectric constant $\epsilon = (1+\sqrt{2})^2$ (curve 1), $\epsilon = 9$ (curve 2). Angle of incidence 30°.

If $N \geqslant 1 + \sqrt{2}$ the reflected wave is circularly polarized for any angle of incidence which satisfies eqn. (8), provided that the metal plate is at a distance given by eqn. (7).

The reflected wave is also circularly polarized when $\tan\frac{1}{2}\delta = -1$, and it is found that, when N and θ have any values such that total reflection occurs, there is always a position of the metal plate for which circular polarization occurs, and the distance d is then given by

$$d = \frac{1}{2k\mu}\log\left\{\frac{N^2 - 1 - \alpha}{N^2 + 1 - (4N^2 + \alpha^2)^{1/2}}\right\}. \qquad \ldots\ldots(9)$$

When $N \geqslant 1 + \sqrt{2}$ there are thus two possible positions of the metal plate at which circular polarization occurs for angles of incidence satisfying eqn. (8), and in these two positions the electric vector rotates in opposite directions. Figure 2 shows the calculated curves for an angle of incidence of 30° and various values of dielectric constant, and shows that the two positions can be obtained for this angle of incidence. It can also be shown that a phase difference of + 90° cannot be obtained no matter how high the dielectric constant, when the angle of incidence is 45°. This is because the inequality in eqn. (8) is never satisfied.

§ 3. EXPERIMENTAL

The klystron valve shown in fig. 3 generates radiation of wavelength about 1·25 cm into a rectangular waveguide, which is then tapered up to form a horn aperture 3 in. square. The horn is 15 cm long, and a polystyrene lens is used to reduce the divergence of the radiated energy. The surfaces of this lens are matched to free space to reduce reflections from them.

A similar horn and lens forms the receiving aperture; the radiation received is then fed into a crystal detector and the rectified current observed on a galvanometer. The horns are mounted on radial slides which can rotate about the centre of a circular base 1 metre in diameter. Since the rectangular waveguide propagates only the H_{10} mode, in which the electric vector is normal to the larger face, this system automatically fixes the polarization of the field which is radiated or received. The supports which carry the horns on the slides are thus provided with a cylindrical graduated head, in which the horns can be rotated about their axes in order to change this polarization.

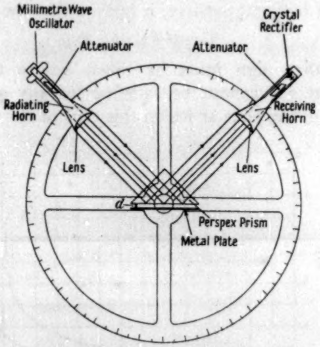


Fig. 3. Arrangement for total reflection experiments.

The totally reflecting surface of the prism is mounted vertically over the centre of the spectrometer, with its faces perpendicular to the axes of the horns, which are set 90° apart on the base scale, and distant 38 cm from the centre. The incident electric field was usually polarized at 45° to the plane of incidence to give equal components perpendicular and parallel to it. Denote the phase change on total reflection between these two components by δ; then it follows that

$$\cos \delta = \frac{P_{\max} - P_{\min}}{P_{\max} + P_{\min}} \qquad \ldots\ldots(10)$$

where $P_{\max}$ and $P_{\min}$ are the maximum and minimum powers received as the receiving horn is rotated about its axis. Note also that provided the amplitudes of the signals remain equal, the axes of the resultant ellipses will always be at 45° and $\frac{1}{2}\pi + 45°$ to the horizontal whatever the phase change.

The measured variation of δ with distance d of the plate is shown in fig. 4, together with the theoretical curve for Perspex for angle of incidence 45°; the agreement between the two curves is quite good. From the experimental curve, it is seen that the phase change δ is $-90°$ when $d = 0{\cdot}084$ cm, and is zero, and the polarization again linear, when $d = 0{\cdot}250$ cm. The theoretical values of d for these two positions are 0·089 cm and 0·243 cm respectively, in close agreement with the experimental values.

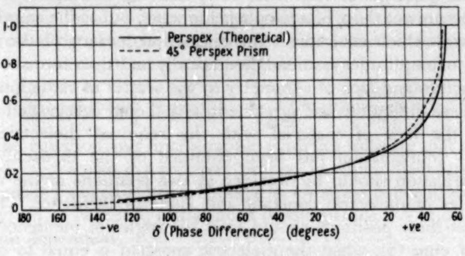


Fig. 4. Effect of metal plate on total reflection. Plot of phase shift δ against distance d of plate from totally reflecting surface, $\lambda = 1{\cdot}25$ cm.

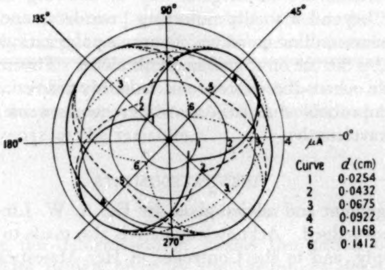


Fig. 5. Effect of metal plate on totally reflecting prism. Plot of received amplitude against rotation of receiving horn about its axis for various distances d of the plate from the surface of the prism.

From the asymptotic value of the phase change δ and d becomes large, and the well-known result for the phase change on total reflection without the metal plate (Stratton 1941), the dielectric constant of Perspex may be deduced. A value of 2·56 is thus found for this, a result which is about 4% below the correct value, but in view of diffraction errors and the scattering of radiation by the prism we may regard the agreement as reasonable.

Figure 5 shows the variation of amplitude with angle of rotation of the receiving horn for different values of d. As d increases from small values the cusp of the curve increases, and the response ideally becomes circular. As d increases beyond this position the curves swing round through 90°, since cos δ

changes sign and the axes are interchanged. The cusp then decreases to zero when $\delta=0$, and then finally increases so that the curve at $d=0{\cdot}971$ cm is approximately that obtained with d infinite. We note also that the axes of the ellipses incident on the receiving horn are all inclined at approximately 45° to the x axis.

§ 4. CONCLUSIONS

The agreement between the theoretical curve, which assumes incident plane waves of equal amplitude, and the experimental curve obtained with apertures 3 in square is quite good. It would seem that the discrepancy occurs because the conditions are not those of a simple plane wave, but of the angular distribution of plane waves which constitutes the diffraction pattern of the aperture field. The equality of amplitudes assumed theoretically may therefore not be maintained, due to diffraction and lack of symmetry in the scatter of radiation by the prism. In view of these difficulties the agreement with theory is regarded as satisfactory.

Due to the lack of material with the required dielectric constant, no experimental confirmation has been obtained of the theoretical result, that circular polarization should be obtained at two positions of the metal plate for suitable values of dielectric constant and angle of incidence. It seems clear, however, that these positions should exist for angles of incidence satisfying the inequality of eqn. (8), when the dielectric constant is equal to or greater than $(1+\sqrt{2})^2$. One way of verifying this would be to use two prisms and two metal plates, when two positions of the plates for circular polarization should be found.

These experiments are an elegant way of showing the presence of the evanescent wave beyond a totally reflecting boundary, and show how it may be utilized to produce millimetre waves of various polarizations, including circular polarized waves, by the use of optical methods alone. This may prove extremely useful in this field where the existence of a doubly refracting material for such purposes seems improbable due to the vast difference between conditions at optical and millimetre wavelengths.

ACKNOWLEDGMENTS

The encouragement and advice given by Dr. A. W. Lines during the work is gratefully acknowledged. Acknowledgment is also made to the Chief Scientist, Ministry of Supply, and to the Controller of Her Majesty's Stationery Office, for permission to publish the paper.

REFERENCE

STRATTON, J. A., 1941, *Electromagnetic Theory* (London : McGraw-Hill), Sec. 9.7.

The Plastic Flow of Iron and Plain Carbon Steels above the A_3-Point

By P. FELTHAM

Physics Laboratories, The British Iron and Steel Research Association, London

Communicated by N. F. Mott; MS. received 9th June 1953

Abstract. The strain rate $d\epsilon/dt$ of the steady flow of iron and plain carbon steels containing from 0·05 to 1·15% carbon, subject to constant stress in the temperature range lying between the A_3-points and the solidus T_s, is shown to depend upon the applied stress σ, the absolute temperature T, the latent heat of melting L, and the volume per atom aw, through the relation

$$\frac{d\epsilon}{dt}=A_0\exp\left[-\frac{nL}{\boldsymbol{k}T}\left(1-\frac{T}{T_s}\right)\right]\sinh\left(\frac{Nnaw\sigma}{4\boldsymbol{k}T}\right).$$

The form of the equation and micrographic evidence indicate that the dominant rate-controlling mechanism is a recovery process in which disordering of groups of n ($\simeq 20$) atoms is induced at grain boundaries by stress concentrations resulting from the piling up of N ($\simeq 50$) dislocations in slip planes. N and n, but not A_0, are composition dependent.

The transient strain ϵ_{tr} is shown to obey the law

$$\epsilon_{tr}=t^m B_0\exp\left[-\frac{H_{tr}}{\boldsymbol{k}T}\left(1-\frac{T}{T_s}\right)\right]\sinh\left(\frac{q_{tr}\sigma}{\boldsymbol{k}T}\right),$$

where m the exponent of the time is approximately equal to $\frac{1}{3}$; B_0 is a composition independent and H_{tr} and q_{tr} are composition dependent constants. A comparison of the last two parameters with corresponding ones in the steady-flow equation suggests that, with certain qualifications, a relation between the steady and the transient flows of the form $\epsilon_{tr}=\text{const.}\,(td\epsilon/dt)^{1/3}$, as derived by Mott, may be valid over a wide range of the variables.

§ 1. INTRODUCTION

In the study of the deformation of metals the period of classifying and collecting of data, which precedes the theoretical development of every science, can be regarded as drawing to a close towards the mid-thirties, when most of the work carried out in that period was summarized in the books *Kristallplastizität, mit besonderer Berücksichtigung der Metalle* by Schmid and Boas, and in *Distortion of Metal Crystals* by Elam, both of which appeared in 1935. At about the same time the basis for the theoretical development of the problem of the deformation of crystalline materials, particularly metals, was laid by the papers of Taylor (1934), Orowan (1934) and Polanyi (1934) on the theory of dislocations, which provided the means for work on the systematization of the experimental material.

Ewald (1936) defined the aim of the ensuing work as "the struggle for the determination of the nature of the parameters", clearly realizing the scientific and technical importance of interpreting the fundamental significance of the constants in the empirically determined laws of deformation.

However, up to the present only limited progress has been made towards the realization of these aims, whether in the case of the stress–strain relation, important at comparatively low temperatures, or the strain–time relation at constant stress which is of particular importance at temperatures above about $0{\cdot}36T_{\text{melt}}$(°K). Thus, in the strain–time relation, which forms the subject matter of the present work, the need for the use of parameters of unknown nature still arises in interpretations of the steady flow of single crystals (Orowan 1934, Kochendörfer 1941, Laurent 1945, Cottrell and Aytekin 1950) and polycrystals (Kanter 1938, Kauzmann 1941, Novick and Machlin 1947, Cottrell and Aytekin 1950, Johnson and Frost 1952) even if the formalism of the theory of dislocations is introduced in interpreting the experimental results. However, while there is similarity between the functional form of the flow equations relating to the steady flow, this cannot be said of the various expressions which have been proposed to account for the first, transient, stage of the flow.

In fact, except for the pioneering work of Andrade (1910, 1914), comparatively little systematic work has been done on the problem of the transient flow in metals and, in the absence of a broad experimental basis, theoretical interpretations of the phenomena were largely speculative. The problem is complex, and few valid generalizations can be arrived at from the study of the available experimental material. Thus, while it is true that in the case of pure polycrystalline metals, and some alloys (Johnson and Frost 1952) at temperatures above about 0·36 of that of their melting points and in a large number of non-metallic and non-crystalline materials (e.g. celluloid, Jessop and Filon 1928) the Andrade $t^{1/3}$ law represents the transient strain very well over a wide range of stresses and temperatures, it does not, with few exceptions (Cottrell and Aytekin 1950, on zinc) appear to hold in the case of polycrystalline metals at temperatures and stresses at which grain boundary slip does not readily occur, or in metal single crystals. The grain boundaries thus appear to play an important role in the transient deformation, immobile boundaries leading to a largely reversible transient deformation characterized by a logarithmic curve, while with boundaries along which slip takes place with comparative ease the curve is of the t^m form with m approximately equal to $\frac{1}{3}$. In the case of single crystals the experimental evidence is particularly difficult to correlate. Firstly, single crystals containing impurities, which would be expected to have a definite yield point, e.g. iron (Gensamer and Mehl 1938), alpha-brass (Burghoff and Mathewson 1941) and impure zinc (Read and Tyndall 1946), show S-shaped transient flow curves, indicating a delay or incubation period prior to the onset of the rapid transient strain rate. In pure lead (Baker *et al.* 1938) and zinc (with some exceptions, ascribed by Cottrell and Aytekin (1950) to work softening), S-shaped curves were not observed. Further, even in the absence of a 'foot' in the curve, the t^m law, with m independent of the stress and the temperature, and with value approximately $\frac{1}{3}$, is not in general obeyed by single crystals, and different workers find varying relations for crystals of the same degree of purity. Thus for zinc of 99·99 and 99·999% purity, flowing between 10 and 60°c, Cottrell and Aytekin (1950) found either S-shaped curves or curves obeying the t^m law, with $m=\frac{1}{3}$, while Weinberg (1952), working with zinc of the same high purity, found m to be 0·6 at room temperature, and Thompson (1952), also working with single crystals of zinc of 99·99% purity, found m values from 0·2 to 3·3 depending upon the temperature, m being equal to 0·5 at room temperature and to 1·0 at about 60°c. The validity of the $t^{1/3}$ law does not appear to be affected by the crystal symmetry of the metal in the case of pure polycrystalline metals (Andrade 1910, 1914); definite relationships between the crystal symmetry and the transient creep behaviour for single crystals do not, however, as yet appear to have been established.

In view of the lack of systematic experimental work on the transient flow of metals, and because of the incomplete understanding of the nature of the parameters characteristic of the metal which appear in the expressions for the steady flow, the problem of the fundamental nature of the rate determining process, or processes, in the flow of metals cannot be regarded as having been solved, a fact which has influenced not only the manner of approach in the present work, but also the formulation of the main objectives. These were primarily:

(*a*) To study the steady flow of pure polycrystalline iron as a function of the temperature and the stress under carefully controlled conditions, chosen so as to minimize complications arising, on the one hand from prior treatments, i.e. from the history of the metal, and on the other hand from chemical reactions or pronounced structural changes such as might arise from progressive work-hardening during the steady flow. Such complications might seriously invalidate the assumption of a mechanical equation of state for the interpretation of the results in the case of the steady flow or lead to pronounced variability of the exponent of time in any power law which might describe the first-stage, transient, flow.

(*b*) To extend this study, under equally simple conditions, to plain carbon steels, thus using the carbon content as a variable in addition to the stress and the temperature.

(*c*) To examine, if possible, the nature of the deformation micrographically.

(*d*) To attempt a representation of the relationship between the strain rate, the stress and the temperature for the steady flow above the A_3-point by means of an expression of a functional form such as had been found generally suitable by other workers in research on the flow of metals at high temperatures, now simplified by the added assumption that progressive work-hardening did not occur.

(*e*) Having determined the constants in such an equation for the pure iron and for the steels, to examine the influence of different carbon contents upon their magnitudes.

(*f*) To search for any relations between the constants, and to discuss, if possible, any theoretical implications of the results.

(*g*) To extend these studies to the transient flow.

(*h*) To examine the relation of the transient flow to the steady flow.

The choice of iron and plain carbon steels as materials, as well as the choice of the high temperature range, above the A_3-points, was determined by the fact that in this range the steels exist as single-phase solid solutions of carbon in gamma iron (austenite), and the pure iron (gamma) and the steels thus represent comparatively simple systems, appropriate in relation to the aims of the work, and also of technical importance. The limits of the range of the observed rates of strain of the steady flow were determined, at the rapid end of the range, by the mechanical and optical limitations of the equipment and the observer, and at the slow end by the desire not to extend individual runs over more than a few hours.

§ 2. EXPERIMENTAL EQUIPMENT AND PROCEDURE

2.1. *Materials*

The materials used in the experiments were spectrographically pure iron (99·99%) in the form of rods 13 cm long, of 0·17 cm diameter, and similarly shaped specimens of 0·23%C, 0·44%C, 0·56%C, 0·79%C basic open-hearth steels, 1·15%C silicon de-oxidized acid open-hearth steel and effervescent, basic 0·05%C open-hearth steel. All the steels used were close in composition to a series of steels for which a number of physical properties, such as the specific heat (up to 1300°C), the thermal expansion (up to 1000°C), and the electrical conductivity (up to 1300°C), had been measured at the National Physical Laboratory (1946). The steels will subsequently be referred to only by reference to their carbon content, as given in the chemical analyses, table 1.

Table 1

	% by weight							
Type of steel	C	Si	Mn	S	P	Cr	Ni	Cu
Rimming. Basic open-hearth	0·05	0·01	0·16	0·04	0·04	0·02	0·06	—
Killed. Basic open-hearth	0·23	0·13	0·60	0·04	0·04	0·09	0·17	0·07
Killed. Basic open-hearth	0·44	0·15	0·74	0·04	0·03	0·07	0·08	0·13
Killed. Basic open-hearth	0·56	0·16	0·61	0·04	0·03	0·04	0·03	0·04
Killed. Basic open-hearth	0·79	0·15	0·54	0·04	0·04	0·03	0·04	0·03
Killed. Acid open-hearth	1·15	0·18	0·35	0·04	0·04	0·05	0·15	—

2.2. *Apparatus*

A diagram of the apparatus is shown in fig. 1. It consists essentially of an Andrade–Chalmers (1932) constant-stress balance made of steel, and two concentric, molybdenum-wound electric furnaces, encased in a continuously evacuated steel chamber. The temperature of the specimen was measured by means of two Pt : Pt-Rh thermocouples inserted from below into the central furnace tube, the thermocouples having their hot junctions at points near the top and the middle of the unstrained specimen respectively. The furnaces, each of length about three times that of the specimen, were differentially wound; the difference in readings between the two thermocouples at temperatures in the experimental range was generally only a few degrees; the mean of the two, which was taken as the specimen temperature, was estimated to be generally within $\pm\frac{1}{2}\%$ of the true actual temperature (°K). The scale was a thin steel ruler, calibrated in divisions of 0·01 in., bent into the arc of a circle concentric with the specimen-bearing arm of the beam, and firmly attached to it. By observing its movement with a low-power microscope provided with a cross-wire, the extension of the specimen was measured during flow. A pan, which could be released whenever desired by means of a hand-operated bellows mechanism, was used for applying the load. The pressure in the vacuum chamber, which was generally of the order of 2μ Hg during runs, was measured by means of a Pirani gauge.

2.3. *Preparation and Mounting of Specimen*

Specimens were prepared from the rods by first covering with sealing wax the ends outside a central portion, measuring 9·8 cm at room temperature and expanding to about 10 cm at the temperatures of the experiments, and then reducing the diameter of the uncovered portions by rapid etching in 40% nitric acid. The wax was subsequently removed, sharp angles were smoothed out on a watchmaker's lathe, and the specimens were mechanically polished. The dimensions of the specimens and the method of fixing them in pin-vices are shown in fig. 2. The inner furnace, within which the specimen was mounted, was lined with a thin steel cylinder smeared with carbon, to form carbon monoxide by reaction with the oxygen in the residual gas in the vacuum chamber. The function of the carbon monoxide was to reduce the oxygen potential of the specimen material, and thus to prevent its oxidation above the A_3-point. This was achieved, for specimens generally appeared bright after removal from the furnace, and chemical analyses of samples carried out on specimens of the 0·05 and 0·79% steels, which had flowed at temperatures up to 1100°c, showed no observable loss of carbon in comparison with specimens of the same steels which had not been in the furnace.

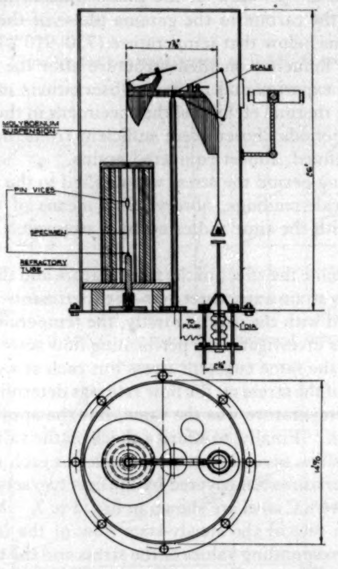

SCALE
MOLYBDENUM SUSPENSION
PIN VICES
SPECIMEN
REFRACTORY TUBE
TO PUMP

Fig. 1. Vacuum chamber, furnace and constant-stress device.

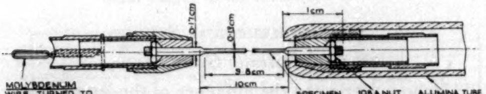

0·17cm
0·13cm
1cm
9·8cm
10cm
MOLYBDENUM WIRE, TURNED TO FIT GROOVE & FORM A LOOP AT THE TOP OF THE PIN VICE
SPECIMEN
10BA NUT
ALUMINA TUBE

Fig. 2. Method of securing specimens.

2.4. *Flow Tests*

In general, the furnace was allowed to warm up as soon as a vacuum of about 2 to $5\,\mu$ Hg was reached. This period extended over about 25 minutes, and a similar time was required to heat the specimen to approximately 1000°c, with an additional 10 to 20 minutes for higher temperatures.

When the testing temperature was reached, the specimen was maintained at that temperature for periods varying from about half-an-hour at 950°c to about five minutes at 1400°c. In view of the phase-transformation at the A_3-point, and the solution of the carbon in the gamma phase of the iron, the mechanical state of the specimens below that temperature (730–910°c) was considered to be without appreciable influence on their structure after the heat treatment at the temperatures in the experimental range. Observations of the austenite grains, made possible by the thermal etching of the specimens in the furnace, showed that the heat-treatment periods chosen were sufficient from the point of view of the formation of well-defined, almost equi-axed grains.

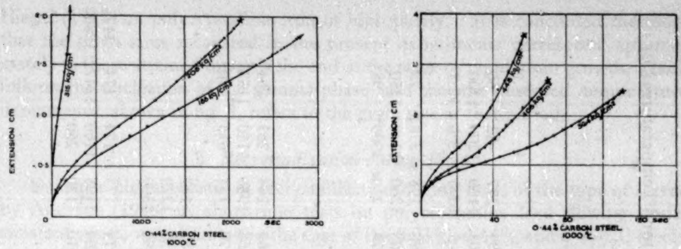
After the annealing period the stress was applied to the specimen by releasing the load, and the scale-readings, observed by means of the microscope, were recorded together with the time indicated by a stopwatch which was started on releasing the load.

In order to determine the effect of the temperature and the applied stress on the steady, quasi-viscous strain rate, three sets of experiments were carried out with each of the steels, and with the iron. Firstly, the temperature dependence of the tensile strain rate was investigated by performing flow tests on a number of specimens flowing under the same constant stress but each at a different temperature. Secondly, the effect of the stress on the flow rate was determined by experiments in each of which the temperature was the same, but the applied stress varied from specimen to specimen. Finally, to afford a check on the validity of any theoretical correlation of the results, several tests were made for each metal at combinations of stresses and temperatures not covered by the first two sets. The flow curves of the representative 0·44%C steel are shown in figs. 3 to 7. Measured values of the natural tensile strain rate of the steady-state flow of the iron and all the steels, together with the corresponding values of the stress and the temperature are given in table 2, and the mean measured grain sizes of the steels are shown in fig. 8. The grain diameters of the pure irons (not shown) were twice that of the 0·23% C steel.

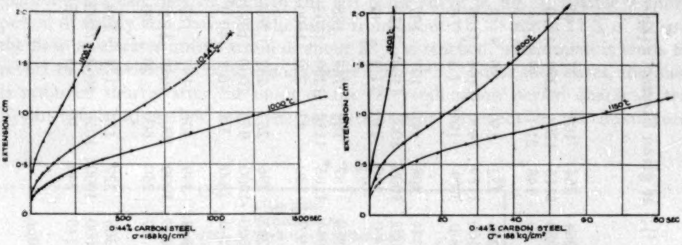
§ 3. METALLURGICAL STUDIES

3.1. *Austenite Grain Size*

Miller (1951) who investigated the influence of the length of the austenitizing time upon the grain size, and Day and Austin (1940) who studied the same problem by means of thermal etching, have shown that the austenite grain size of plain carbon steels does not remain constant at a given temperature, but increases continuously as the period of heating is prolonged. Experiments carried out by Cole (1953) on 0·78 and 0·38%C steels maintained at 850°c *in vacuo*, showed that austenite grain growth was very pronounced in the first 15 minutes, but then rapidly decelerated, until about 5 minutes later growth took place at such a slow rate that no further appreciable increases in grain size took place during the following half hour. Such a behaviour is qualitatively in agreement with the results of Miller, Day and Austin, as well as with those obtained by Cleaves and



Figs. 3 and 4. Flow under constant stress of 0·44%C steel at 1000°c.



Figs. 5 and 6. Flow of 0·44%C steel under a constant stress of 188 kg cm^{-2}.

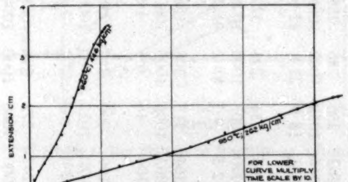


Fig. 7. Flow of 0·44%C steel under constant stress.
Initial length of all specimens 10 cm.

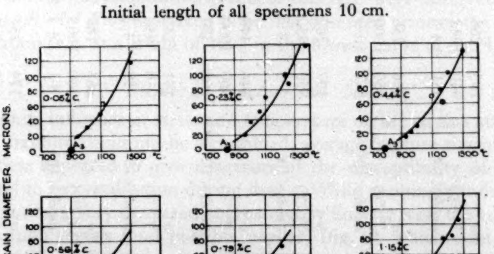


Fig. 8. Temperature dependence of grain-size in plain carbon steels above the A_3-point. Open circles from N. T. Gudtsov *et al.* (1950), crossed circles from L. I. Shushpanov (1937).

Table 2. Measured Values of the Steady Strain Rate $d\epsilon/dt$ (sec^{-1}); at given Stresses σ (kg cm^{-2}) and Temperatures (°c)

Steel (%C)																		
0·00	σ	195	264	300	330	264	264	233	264	264	103	77	40					
	T (°c)	940	942	940	943	955	1000	1050	1080	1115	1204	1270	1370					
	$d\epsilon/dt \times 10^5$	1·1	4·6	9·6	17·6	7·3	20·8	24·0	69	148	21·1	39	—					
0·05	σ	246	256	248	262	256	256	256	256	252	256	80	158	199	277	82	66	
	T (°c)	950	1000	1040	1045	1058	1062	1066	1100	1159	1400	1104	1106	1105	1100	1200	1300	
	$d\epsilon/dt \times 10^5$	3·7	10·9	12·8	18·0	18·5	23·8	22·0	38·0	78·4	$>10^2$*	1·3	7·5	16·5	61·0	5·0	12·3	
0·23	σ	197	262	254	195	254	294	322	384	408	441	197	197	198	198	238	73	
	T (°c)	950	950	958	1000	1005	1000	1006	1002	1000	1000	1100	1150	1208	1260	1300	1350	
	$d\epsilon/dt \times 10^5$	1·5	2·7	2·7	2·5	9·3	18·2	45·2	84·0	140	240	21·0	43	174	230	1000*	98	
0·44	σ	446	448	405	364	315	208	188	188	187	188	188	188					
	T (°c)	960	1000	1000	1003	1005	1000	1000	1044	1105	1150	1200	1303					
	$d\epsilon/dt \times 10^5$	168	380	180	104	46	6·5	4·6	9·7	30	83	220	1500					
0·56	σ	141	242	171	199	200	212	238	282	296	262	262	262	262	264	263	264	
	T (°c)	950	950	1000	1020	1022	1000	1000	1000	1000	950	1000	1050	1058	1086	1150	1153	
	$d\epsilon/dt \times 10^5$	59	4·2	2·5	7·4	8·9	7·2	9·8	25·0	28·0	5·8	16·3	40·0	70	131	615	470	
0·79	σ	263	187	212	242	273	298	326	353	399	436	187	187	187	187			
	T (°c)	960	1005	1000	1000	1000	1000	1000	1000	1000	1005	1050	1100	1150	1220			
	$d\epsilon/dt \times 10^5$	8·1	6·3	9·8	16	36	29	89	57	250	600	26	100	290	1500			
1·15	σ	174	251	174	198	220	254	297	306	327	174	174	139	174	173	174	174	117
	T (°c)	950	954	1000	1000	1000	1005	1000	1000	1000	1050	1050	1090	1104	1105	1166	1247	1280
	$d\epsilon/dt \times 10^5$	9·5	45	25	33	53	81	170	140	240	59	45	35	110	120	260	—	—

* Approximate value.

Hiegel (1942) on polycrystalline iron of high purity. It is concluded therefore that the grain sizes measured in the present experiments correspond approximately to those attained towards the end of the stage of rapid grain growth, which follows the nucleation of the gamma-phase, and that the observed temperature dependence, shown in fig. 8, refers to the grain size at that period.

3.2. *Recrystallization during Flow*

Evidence characteristic of recrystallization during flow, of the type observed by Andrade (1948) in short-time tests on polycrystalline lead flowing under constant stress, was obtained in the case of the 0·23%, 0·44% and 0·56%C steels, particularly at high levels of stress (figs. 4 and 7). The principal features of the phenomenon can best be seen in the left-hand curve in fig. 7. After a short period of steady flow, covering the range from about 8% to about 11% of strain, the flow accelerates until a strain of about 28% is reached, whereupon it tends to revert to the original steady, second-stage flow rate. If, in such cases, the load is removed shortly after the onset of the recrystallization period characterized by an accelerating flow rate, the recrystallization proceeds in the unstressed specimen, and on re-applying the load after about 10 minutes a strain rate corresponding closely to the initial, steady strain rate is maintained until fracture takes place. Whenever recrystallization effects of this type were observed, the steady strain rate $d\epsilon/dt = l^{-1}\,dl/dt$ was taken to be that observed prior to the onset of the recrystallization (e.g. at a strain of 10% in the lowest curve of fig. 4).

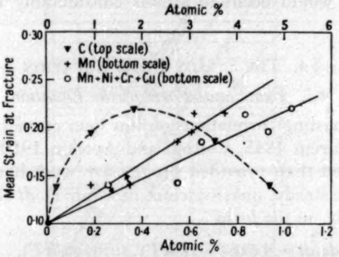


Fig. 9. Mean natural strain at the instant of fracture as a function of composition.

3.3. *Strains at Fracture and 'Necking'*

As a definite influence of stress and temperature on the strains attained up to the time of fracture could not be ascertained, average fracture strains were computed. These appeared to give a measure of the susceptibility of the various types of steel to recrystalization during flow. While seemingly independent of the carbon content, they increased approximately linearly with the concentration of the transition metals (and possibly copper) (fig. 9), which exist in the iron predominantly in substitutional solid solution, and have Goldschmidt radii within 4% of the former.

In general, specimens deformed fairly uniformly during flow, up to strains of about 15%. Subsequently 'necking' frequently took place, as was confirmed by measuring the dimensions of the specimens after flow by means of a micrometer. Sections of the flow curves (figs. 3–7), which could almost certainly be associated with 'necking' are shown as broken lines.

3.4. *Micrographic Observations*

The most typical features observed in microscopic examinations of the austenite grain structure on the surfaces of specimens, after flow, were relative displacements of grains with respect to one another, as disclosed by discontinuities in fiducial scratch-marks at grain boundaries; the coarse appearance of boundaries oriented approximately at right angles to the tensile axis, both of which effects are visible in Plate I, and the occurrence of curvature in intragranular segments of scratch-marks which were originally straight (Plate II).

With the exception of the apparent grain-boundary coarsening, the origin of which could not be satisfactorily accounted for, the discontinuities and the curving of scratch marks provided definite evidence of both intergranular and intragranular deformation. In view of the large strains observed, the occurrence of both types of deformation is in fact to be anticipated, for if only boundary movements were responsible for the flow, intergranular hole-formation and subsequent fracture would occur at strains considerably below those actually attained.

§ 4. THE STEADY-FLOW EQUATION

4.1. *Functional Form of the Equation*

On the basis of existing correlations of flow data on metal single crystals and polycrystals (e.g. Laurent 1945, Cottrell and Aytekin 1950, Johnson and Frost 1952), it was assumed that, provided progressive work-hardening did not take place during flow, the steady, quasi-viscous, flow rate $d\epsilon/dt$ should be expressible, at least approximately, in the form

$$d\epsilon/dt = A \exp(-H/\boldsymbol{k}T) \,.\, \sinh(q\sigma/\boldsymbol{k}T), \qquad \ldots\ldots (1)$$

where σ is the applied tensile stress, T the absolute temperature, and A, H and q three constants characteristic of the type of steel. With resolutions of about one micron, no slip-bands were observed in the surface of the specimens after flow, so that the assumption relating to the absence of progressive work-hardening appeared to be justified.

For stresses high enough (over about $100\,\text{kg}\,\text{cm}^{-2}$) to render possible the replacement of $\sinh(q\sigma/\boldsymbol{k}T)$ by $\frac{1}{2}\exp(q\sigma/\boldsymbol{k}T)$, it was in fact found that, at a fixed temperature, a linear relation was obtained between $\log(d\epsilon/dt)$ and σ for the iron and all steels, in conformity with the requirements of eqn. (1). The numerical values of q, obtained from the slope of these lines, are given in table 3. In a similar manner, by keeping the stress invariant at a value high enough to allow the replacement of $\sinh(q\sigma/\boldsymbol{k}T)$ by $\frac{1}{2}\exp(q\sigma/\boldsymbol{k}T)$, $-H+q\sigma$ can be determined, and thus H found. Alternatively, the value of H can be obtained in a more general way, including all experimental points, from the slope of the line relating $\ln[(d\epsilon/dt)/\sinh(q\sigma/\boldsymbol{k}T)]$ to T^{-1}, again using values of q as determined by the method described, in which case it is not necessary to restrict the stress to a constant, high, value. The application of the second method is shown in fig. 10, and table 3 gives the values of H thus derived, together with the corresponding values of A obtained by substituting in eqn. (1) values of H, q, σ and corresponding, smoothed, values of $d\epsilon/dt$ obtained from the linear relationships (not shown) between $\log(d\epsilon/dt)$ and σ, for $T=\text{const}$.

Table 3

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
0·00	50·0	73·6	$8{\cdot}5\times10^{6}$	1808	20·0	6·71	6·93
0·05	49·8	61·2	$3{\cdot}7\times10^{4}$	1800	16·9	5·00	4·57
0·23	43·6	73·6	$7{\cdot}2\times10^{6}$	1753	20·0	6·99	6·86
0·44	43·2	84·8	$9{\cdot}6\times10^{8}$	1705	23·3	8·96	8·98
0·56	40·4	94·2	$6{\cdot}2\times10^{10}$	1673	26·3	10·70	10·79
0·79	41·8	102·6	$1{\cdot}6\times10^{12}$	1620	28·1	12·40	12·20
1·15	38·2	61·4	$1{\cdot}0\times10^{6}$	1533	16·9	6·52	6·00

(1) Steel (%C); (2) q ([cal/g atom]/kg cm^{-2}); (3) H (kcal/g atom); (4) A (sec^{-1}); (5) T_s (°K); (6) $n=H/L$; (7) $\log_{10}[2{\cdot}75\times10^4\exp(1{\cdot}16H/\boldsymbol{k}T_s)]$; (8) $\log_{10}A$.

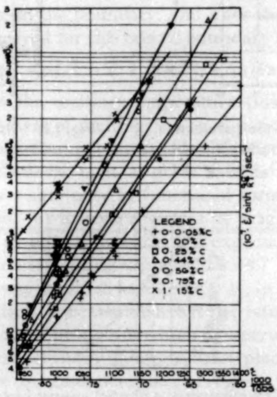


Fig. 10. Interpretation of the steady flow by means of the relation $d\epsilon/dt=A\exp(-H/\boldsymbol{k}T)\,.\,\sinh(q\sigma/\boldsymbol{k}T)$.

4.2. *Relations between Parameters*

The use of the carbon content as a variable made possible the discovery of the relations between H and $\log A$ (fig. 11), H and %C (fig. 12) and q and %C (fig. 13).

It can be seen from the figures that the $(H, \log A)$ relationship applies to the iron and all the steels, only the point corresponding to the 1·15%C steel lying somewhat off the line. This relationship must therefore be regarded as more fundamental than that between H and %C which, as can be seen from table 3 and fig. 12, holds only for the basic carbon steels but not for the iron or the 1·15% acid steel. Consequently the carbon content can be regarded as of primary significance from the point of view of the magnitudes of H, if the steels, carbon content apart, are similar as far as the state and the quantity of other impurities are concerned. The relation between q and %C (fig. 13) appears to conform to the equation

$$q = \alpha \exp(-\beta . \%\mathrm{C}), \qquad \ldots\ldots(2)$$

where α and β are positive constants, but the scatter of the points is considerable, and eqn. (2) can be regarded only as an approximate expression of a general tendency. Since the relation between H and $\log A$ was indicative of a temperature dependent activation energy, the ratio H/T_s, where T_s is the solidus temperature (table 3), i.e. the temperature at which the grain boundaries melt, was plotted against $\log A$. This relation was again found to be linear to about the same degrees of accuracy as the $(H, \log A)$ relationship. The equation of the line was found to be

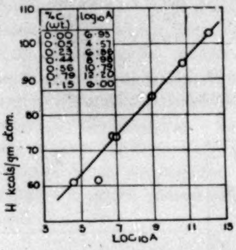


Fig. 11. Relation between the parameters A and H appearing in the flow equation $d\epsilon/dt = A \exp(-H/\boldsymbol{k}T) . \sinh(q\sigma/\boldsymbol{k}T)$.

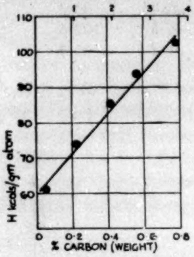


Fig. 12. Relation between H and the carbon content in hypo-eutectoid steels.

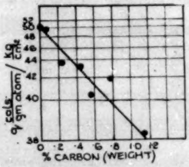


Fig. 13. The parameter q plotted against carbon content.

$$A = A_0 \exp(rH/\boldsymbol{k}T_s), \qquad \ldots\ldots(3)$$

with A_0 now independent of the type of steel. The actual values were

$$A = 2{\cdot}75 \times 10^{-4} \exp(1{\cdot}16H/\boldsymbol{k}T_s)\,\mathrm{sec}^{-1}, \qquad \ldots\ldots(4)$$

so that eqn. (1) could now be written in the form

$$\frac{d\epsilon}{dt} = A_0 \exp\left[-\frac{H}{\boldsymbol{k}T}\left(1 - 1{\cdot}16\frac{T}{T_s}\right)\right] \sinh\left(\frac{q\sigma}{\boldsymbol{k}T}\right). \qquad \ldots\ldots(5)$$

The values of $\log_{10}[2{\cdot}75 \times 10^{-4} \exp(1{\cdot}16H/\boldsymbol{k}T_s)]$ are shown, side by side with the corresponding, experimental, values of $\log_{10} A$ in table 3. If it is considered that, apart from other approximations probably implied in the adoption of eqn. (1), (*a*) the assumption that A is independent of the temperature and the grain size is in all probability only a very rough approximation, (*b*) that the values of T_s employed were those of the unstressed, pure binary Fe/C system, and (*c*) that experimental errors affect the numerical values of constants r and A_0, it is not unreasonable to assume that if appropriate allowance could be made for the sources of error listed above, the activation energy, apart from the stress-dependent part exemplified by the sinh-terms, would take the form $H(1-T/T_s)$.

4.3. *The Latent Heat of Melting as a Parameter in the Flow Equation*

In dealing with the problem of slip at grain boundaries, Mott (1948) assumed that if two crystals are in contact, but owing to differences in orientation cannot fit, the surface of contact could be sub-divided into islands where fit is reasonably good, separated by lines near which fit is bad. Considering one such island, Mott supposed that, say, the top plane could move under the influence of a shear stress relatively to the one underneath from one equilibrium position to another, and that this was the elementary act responsible for intercrystalline slip. If each island is assumed to contain n atoms and to cover an area nw then, writing τ for the shear stress acting on the boundary, and a for the interatomic spacing in the shear direction, he derived for the rate of boundary slip v the expression:

$$v=2\nu a\exp\left[-nL(1-T/T_m)/\boldsymbol{k}T\right].\sinh\left(\tau nwa/2\boldsymbol{k}T\right), \quad \ldots\ldots(6)$$

where $\nu\simeq 10^{13}$ sec, T_m is the melting temperature (identical with T_s in pure metals), and L the latent heat of melting. The equation is reconcilable with a model in which clusters of n atoms (estimated by Mott at about 14 in aluminium) would re-arrange themselves, i.e. melt, at grain boundaries.

Although there exists a similarity in functional form between eqns. (5) and (6), the terms $2\nu a$ in eqn. (6) is several orders of magnitude greater than the corresponding term A_0 in eqn. (5), while the expression $\tau nwa/2\boldsymbol{k}T$ is about fifty times less than the corresponding expression $q\sigma/\boldsymbol{k}T$ in eqn. (5), if τ is taken to be $\frac{1}{2}\sigma$. The equation was modified as follows. If, in view of the evidence of intra-crystalline deformation, it is assumed that the rate-controlling mechanism in the steady flow is a recovery process in which co-operative phenomena equivalent to remelting of small clusters of atoms take place, particularly at grain boundaries, and that these processes take place in regions where a stress concentration exists, such as would arise from the piling up of dislocations in slip-planes at grain boundaries, it should take the form

$$d\epsilon/dt=A_0\exp\left[-nL(1-T/T_s)/\boldsymbol{k}T\right].\sinh\left(Nnwa\sigma/4\boldsymbol{k}T\right), \quad \ldots\ldots(7)$$

where N is numerically equal to the number of dislocations piled up at a barrier and where A_0 would be expected to have a lower value than $2\nu a$. In the case of iron ($n=20$, table 3), remembering that τ has been replaced by $\frac{1}{2}\sigma$, the relation

$$q=\tfrac{1}{4}Nnwa \quad \ldots\ldots(8)$$

then yields, on writing $wa=11{\cdot}8\times10^{-24}$ cm^3 and $q=3{\cdot}44\times10^{-21}$ cm^3($=50{\cdot}0$ cal cm^2 (gm atom)$^{-1}$ kg^{-1}) in eqn. (8), $3{\cdot}44\times10^{-21}=\frac{1}{4}\times20\times11{\cdot}8\times10^{-24}N$, giving

$$N=59, \quad \ldots\ldots(9)$$

which is a reasonable value, similar to that estimated by McLean (1952) on the basis of his work on aluminium, in which he found 'fine' slip, not resolvable by ordinary optical means, to be the principal source of deformation. Grain-boundary slip also observed by him appears to be dependent on the occurrence of intragranular slip, an observation which is in agreement with the interpretation of the rate-controlling mechanism of the steady flow in the present work. The highest value of N (eqn. (9)) was found to be 67 (0·05%C steel), and the lowest 35 (0·79%C).

§ 5. The Transient-Flow Equation

5.1. *Functional Form of the Equation*

In fig. 14 strains of the transient portions of the flow curves of the 0·44%C steel, shown in figs. 3 to 7, are plotted against $t^{1/3}$. The linear relationship obtained confirms the validity of the $t^{1/3}$ law. Because of the uncertainties in the measured magnitudes of the initial extensions, occurring within the first one or two seconds—uncertainties derived particularly from the unavoidable slight slackness in the molybdenum suspension wire—the curves were drawn (by minor displacements parallel to the strain axis) so as to pass through the origin when extrapolated backwards. The errors thereby introduced are small, since only small displacements were generally required to effect this adjustment. The intercepts, on the strain axis, at unit time, of the extrapolated straight lines give the numerical value of the parameter β in the expression $\epsilon_{tr} = \beta t^{1/3}$, which is the equation of the lines in fig. 14. The stresses, temperatures and β-values corresponding to the individual lines are given in table 4. Strain–time plots on a log–log scale showed that the exponent of the time was, in general, 10–20% higher than $\frac{1}{3}$, and that the exponent was not entirely independent of the stress. Thus, for example, it increased from about

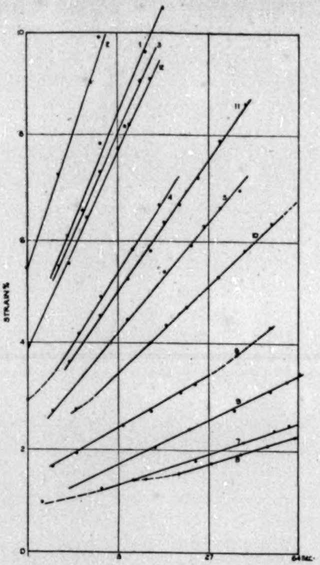


Fig. 14. Relation between the strain and the time for 0·44%C steel. The time is plotted on a cube-root scale. For legend see table 4.

0·3 to 0·4 in the range 100–450 kg cm^{-2} at 1000°c for the 0·44% C steel, and fell from about 0·6 at 100 kg cm^{-2} to about 0·4 at 400 kg cm^{-2} at 1000°c for the 0·79% C steel. Such large variations were not observed however, except for the 0·79% C steel. Nevertheless, for strains after about 10 seconds the curves for the 0·79% C steel could be represented quite well by means of the $t^{1/3}$ law, and this was used in this case as well to determine the β-values. The exponent also appeared to depend somewhat upon the temperature, but this could not be established with certainty.

Table 4. Transient Flow of 0·44%C Steel

(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	960	446	4·05	28·5	43·5	1·5
2	1000	448	3·94	25·7	53·5	2·1
3	1000	405	3·57	17·5	40·1	2·3
4	1000	364	3·20	12·2	26·7	2·2
5	1000	315	2·77	7·9	20·4	2·6
6	1000	264	2·32	4·8	11·6	2·4
7	1000	208	1·83	3·0	6·1	2·0
8	1000	188	1·66	2·5	5·7	2·3
9	1044	188	1·54	2·3	8·4	3·7
10	1105	188	1·52	2·2	17·1	7·8
11	1150	188	1·48	2·1	25·2	12·0
12	1200	188	1·43	2·0	37·4	18·7
13	1303	188	—	—	—	—

(1) Specimen; (2) Temperature (°c); (3) σ(kg cm^{-2}); (4) $q_{\mathrm{tr}}\sigma/\boldsymbol{k}T$; (5) sinh $(q_{\mathrm{tr}}\sigma/\boldsymbol{k}T)$; (6) $10^3\beta$ (sec$^{-1/3}$); (7) $10^3\beta/\sinh(q_{\mathrm{tr}}\sigma/\boldsymbol{k}T)$.

The β-values from the first set of experiments, in which the temperature was kept constant and the stress varied, were plotted logarithmically against the stress for all the steels (fig. 15), but not for the iron, for which the scatter of the results was too large. The linear relationship obtained, together with the observed tendency of the line of lowest slope (0·23% C, fig. 15) to curve towards the stress axis for low values of σ, and the observations made by Andrade (1910, 1914) that the (β, σ) curve (T = constant) passes through the point $\beta = 0$, $\sigma = 0$, show that at a given temperature β is proportional to sinh$(r\sigma)$, where r is a constant. This relation in turn suggests a stress dependence of β of the form sinh$(q_{\mathrm{tr}}\sigma/\boldsymbol{k}T)$ frequently found in stress-dependent rate processes as a result of the influence of

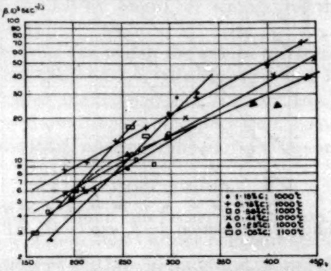


Fig. 15. Stress dependence of β at constant temperature (1000°c).

the stress on the activation energy. The values of q_{tr} were determined from the slopes of the $[\beta, \sinh(q_{tr}\sigma/\boldsymbol{k}T)]$ relationship, and are given in table 5.

Table 5

Steel (%C)	β_0 (sec$^{-1/3}$)	H_{tr}	H	$\dfrac{H_{tr}}{H}$	q_{tr}	q_1	$\dfrac{q_{tr}}{q}$	T_s (°K)	$\dfrac{H_{tr}}{\boldsymbol{k}T_s}$	$\ln\beta_0 + 2{\cdot}5$
0·00	$1{\cdot}6\times10^3$	38·6	73·6	0·53	36·0	50·0	0·72	1808	10·7	9·9
0·05	$1{\cdot}9\times10^6$	61·0	61·2	1·00	46·0	49·6	0·94	1800	16·9	16·9
0·23	$8{\cdot}1\times10^2$	30·8	73·6	0·42	16·6	43·6	0·38	1753	8·8	9·0
0·44	$7{\cdot}4\times10^3$	38·3	84·8	0·45	22·4	43·2	0·53	1705	11·2	10·9
0·56	$3{\cdot}1\times10^3$	37·2	94·2	0·40	25·0	40·4	0·61	1673	11·1	10·4
0·79	$8{\cdot}3\times10^5$	49·2	102·6	0·48	22·0	41·8	0·53	1620	15·2	16·1
1·15	$3{\cdot}5\times10^6$	59·0	61·4	0·97	42·6	38·2	1·11	1533	19·3	17·5

H in k cal/g atom, q in (cal/g atom) kg cm^{-2}.

Thus, if the assumption of a stress-dependent activation was correct, it was to be expected that the relation

$$\beta = \beta_0 \exp(-H_{tr}/\boldsymbol{k}T) \,.\, \sinh(q_{tr}\sigma/\boldsymbol{k}T) \qquad \ldots\ldots(10)$$

would hold, β_0 now being a constant independent of the stress σ and the temperature T, with a characteristic value for each steel. In that case a graph of $\log[\beta/\sinh(q_{tr}\sigma/\boldsymbol{k}T)]$ against T^{-1} should be linear, and the values of the activation energies H_{tr} should be obtainable from the slope. The validity of the hypothesis was established, as can be seen by reference to fig. 16. The values of H_{tr} determined by means of fig. 16 are given in table 5. It follows therefore that the transient strain ϵ_{tr} of iron and of plain carbon steels containing from 0·05 to 1·15% C, flowing under constant stress in the temperature range 950–1400°c, and at stresses leading to steady strain rates 10^{-5} to 10^{-2} per second, can reasonably well be represented by the rate process equation:

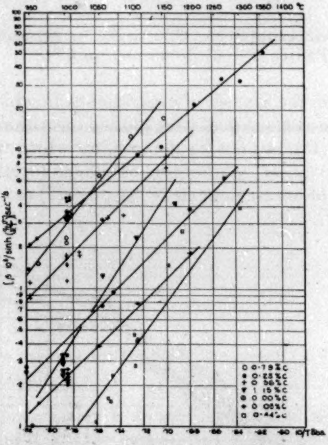


Fig. 16. Validity of the relation $\beta = \beta_0 \exp(-H_{tr}/\boldsymbol{k}T) \,.\, \sinh(q_{tr}\sigma/\boldsymbol{k}T)$.

$$\epsilon_{tr} = t^{1/3}\beta_0 \exp(-H_{tr}/\boldsymbol{k}T) \,.\, \sinh(q_{tr}\sigma/\boldsymbol{k}T), \qquad \ldots\ldots(11)$$

in which, within the range of variables investigated, β_0, H_{tr} and q_{tr} can be regarded as depending only upon the composition of the steels.

5.2. *Relation between Parameters*

A definite relation appeared to exist between the values of H_{tr} and the corresponding values of the parameters β_0 for the iron and all the steels investigated. It was found to be:

$$\ln \beta_0 + 2{\cdot}5 = H_{tr}/\boldsymbol{k}T_s, \qquad \ldots\ldots(12)$$

so that the transient strain could be expressed in the form

$$\epsilon_{tr} = t^{1/3} B_0 \exp\left[-\frac{H_{tr}}{\boldsymbol{k}T}\left(1-\frac{T}{T_s}\right)\right] . \sinh(q_{tr}\sigma/\boldsymbol{k}T), \qquad \ldots\ldots(13)$$

with B_0 now independent of the stress, temperature and the composition of the steels. The value of B_0 was found to be $8{\cdot}1 \times 10^{-2}\,\mathrm{sec}^{-1/3}$ (i.e. $\exp(-2{\cdot}5)$). In table 5 the values of $\ln \beta_0 + 2{\cdot}5$ and corresponding values of $H_{tr}/\boldsymbol{k}T$ are shown side by side for the iron and steels used; experimentally determined values of β_0 and H_{tr} having been substituted. The agreement can be seen to be satisfactory. The activation energy of the rate process involved in the transient flow thus appears to fall to zero at the solidus, i.e. at the point at which melting begins to take place at grain boundaries.

§ 6. THE RELATION BETWEEN THE TRANSIENT AND STEADY FLOWS

A striking similarity can be seen to exist between the equation for the strain rate of the steady flow (eqn. (7) or (5)) and eqn. (13), which, as the ratios of the corresponding parameters H_{tr} and H, and q_{tr} and q (table 5) indicate, appears to have a definite significance. Mott (1953) recently developed a fairly general theory of the transient flow which leads to the expression

$$\epsilon_{tr} = \mathrm{const}\,(td\epsilon/dt)^m, \qquad \ldots\ldots(14)$$

with $\frac{1}{3}$ as the most likely value for the exponent m, in the case of pure metals. The validity of eqn. (14) in the present case would then imply (replacing the sinh-term in eqns. (7) and (13) by $\frac{1}{2}$ exp, for simplicity) that $H_{tr}/H = q_{tr}/q = m$, with m about $\frac{1}{3}$ (see table 5).

With the exception of iron, for which the values of H_{tr} and q_{tr} could not be determined accurately, and the 0·05%C (effervescent) and 1·15%C (acid) steels, of which the latter showed an anomalous, low, H value, eqn. (14) thus appears to describe the transient flow above the A_3-point quite well, particularly if it is recalled that m, the exponent of the time, was generally found to be somewhat in excess of $\frac{1}{3}$.

§ 7. CONCLUSIONS

The principal conclusions which can be arrived at from the present study of the flow of iron and plain carbon steels under constant stress in the temperature range lying between the solidus and the A_3-points, at stresses leading to strain rates of the steady flow ranging from 10^{-5} to 10^{-2} per second, are as follows:

(*a*) Within the limits of experimental error, the stress and temperature dependence of the steady, second-stage (quasi-viscous) flow-rate $d\epsilon/dt$ in both pure iron and plain carbon steels can be expressed by means of an equation containing three parameters, A, H and q, which as a first approximation are taken to be constant for a given steel. The equation takes the form

$$d\epsilon/dt = A \exp(-H/\boldsymbol{k}T)\sinh(q\sigma/\boldsymbol{k}T),$$

indicative of a rate process.

(*b*) The occurrence of strains of 20% or more, together with micrographic evidence confirms that intracrystalline and intercrystalline deformations take place, of which the former is ascribed to fine slip such as had been shown to occur at elevated temperatures in recent papers on the flow of metals.

(*c*) The value of the activation energy H increases with the carbon content, in the case of otherwise similar basic carbon steels, from 61·2 kcal/g atom for a 0·05% C steel to 102·6 kcal/g atom for a 0·79% C steel, and is thus not identifiable with the activation energy for self-diffusion which is known to decrease with carbon content from about 68 kcal/g atom for ingot iron to 33 kcal/g atom for a 1·06% C steel (Gruzin *et al.* 1951) within the same range of temperatures.

(*d*) The value of H is strongly dependent upon the type of steel; the linear relation between H and %C described in (*c*) cannot be extrapolated to pure iron or the 1·15%C steel which was of the acid type.

(*e*) While, in view of the high structure sensitivity, H cannot be regarded as a fundamental physical constant, the linear relation found to hold between $\log_{10} A$ and H applies to the iron and all the steels investigated, indicating a temperature-dependent activation energy which, apart from the stress dependence accounted for by the sinh-term, takes the form $H(1-T/T_s)$.

(*f*) This result taken in conjunction with the experimentally determined values of q is consistent with a rate-determining mechanism in which remelting of small groups of atoms, induced by stress concentrations resulting from the piling-up of dislocations, takes place at obstacles such as grain boundaries. The magnitude of the stress concentration factor N, and the size of the groups n, are of the order of 35 to 67 times and 17 to 28 iron atoms respectively, the actual values depending in a definite manner upon the amount of carbon and also upon the nature and distribution of other impurities.

(*g*) On introducing the stress concentration N, the number of iron atoms in a remelting group n, the latent heat of melting L, and the parameter A_0 which is common to the iron and all the steels investigated, the steady natural strain rate can be expressed by the equation

$$\frac{d\epsilon}{dt} = A_0 \exp\left[-\frac{nL}{\boldsymbol{k}T}\left(1-\frac{T}{T_s}\right)\right] \sinh\,(Nnaw\sigma/4\boldsymbol{k}T)$$

where T_s is the solidus temperature and aw the volume per atom.

(*h*) In so far as recovery at grain boundaries thus appears to be rate-determining it is highly probable that A_0 will depend somewhat upon the grain size, although this was not quantitatively established.

(*i*) The transient strain can be expressed with fair accuracy by a power law of the Andrade type, $\epsilon_{tr} = \beta t^m$, where the exponent of the time takes values of the order $\frac{1}{3}$, most frequently about 10–20% higher.

(*j*) The exponent m appears to be somewhat stress dependent, increasing with stress in a given range of stress in some steels, and decreasing in others.

(*k*) The exponent m appears to be slightly temperature dependent, but this could not be definitely established.

(*l*) The stress and temperature dependence of the parameter β can be expressed in the form

$$\beta = B_0 \exp\left[-\frac{H_{tr}}{\boldsymbol{k}T}\left(1-\frac{T}{T_s}\right)\right] \sinh\,(q_{tr}\sigma/\boldsymbol{k}T).$$

(*m*) With some exceptions the transient strain ϵ_{tr} obeyed the equation $\epsilon_{tr} = \text{const}\,(t\,d\epsilon/dt)^m$, proposed by Mott, quite well.

(*n*) In consequence of the approximate validity of the equation in (*m*), the assumption made in deriving it, namely that the fundamental mechanism operative in the transient and steady flows at temperatures above about $0{\cdot}36\,T_{\text{melt}}$ (°K) are similar in essence, gains plausibility, and

(*o*) the decelerating nature of the transient flow must therefore be regarded as arising primarily from the diminution with time of the frequency of occurrence of the unit recovery process, the frequency falling to a steady value characteristic of the steady, quasi-viscous, flow.

ACKNOWLEDGMENTS

The author would like to express his gratitude to Professor N. F. Mott for his encouragement and continued interest in the research. The discussions with him, on several occasions in the period in which the work was carried out, have been an invaluable source of guidance and inspiration.

REFERENCES

ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc.* A, **84,** 1; 1914, *Ibid.*, **90,** 329; 1948, *Nature, Lond.*, **162,** 410.
ANDRADE, E. N. DA C., and CHALMERS, B., 1932, *Proc. Roy. Soc.* A, **138,** 348.
BAKER, J. B., BETTY, B. B., and MOORE, H. F., 1938, *Metals Tech.*, **5,** 1.
BURGHOFF, H. L., and MATHEWSON, C. H., 1941, *Trans. Amer. Inst. Min. Met. Engrs.*, **143,** 45.
CLEAVES, H. E., and HIEGEL, J. M., 1942, *Bur. Stand. J. Res., Wash.*, **28,** 643.
COLE, D. G., 1953, *J. Iron Steel Inst.*, in the press.
COTTRELL, A. H., and AYTEKIN, V., 1950, *J. Inst. Met.*, **77,** 389.
DAY, M. J., and AUSTIN, J. B., 1940, *Trans. Amer. Soc. Met.*, **28,** 354.
EWALD, P. P., 1936, *Naturwissenschaften*, **24,** 277.
FELTHAM, P., 1950, *Metallurgia*, **4,** 60.
GENSAMER, M., and MEHL, R. F., 1938, *Trans. Amer. Inst. Min. Met. Engrs.*, **131,** 372.
GRUZIN, P. L., KORNEV, Y. V., and KURDYUMOV, G. V., 1951, *Dok. Akad. Nauk S.S.S.R.*, **80** (1), 49.
GUDTSOV, N. T., LOZINSKII, M. G., ZUDIN, F. I., BOGDANOV, N. A., and MATVEEVA, M. P., 1950, *Izv. Akad. Nauk S.S.S.R.*, O.T.N., (1), 108.
JESSOP, T. H., and FILON, L. N. G., 1928, *Phil. Trans. Roy. Soc.* A, **223,** 89.
JOHNSON, A. E., and FROST, N. E., 1952, *J. Inst. Met.*, **81,** 93.
KANTER, J. J., 1938, *Trans. Amer. Inst. Met. Eng.*, **131,** 385.
KAUZMANN, W., 1941, *Trans. Amer. Inst. Met. Eng.*, **143,** 57.
KOCHENDÖRFER, A., 1941, *Plastische Eigenschaften von Kristallen* (Berlin : Springer).
LAURENT, P., 1945, *Rev. Métall.*, **42,** 125.
MCLEAN, D., 1952, *J. Inst. Met.*, **80,** 507, **81,** 133.
MILLER, O. O., 1951, *Trans. Amer. Soc. Met.*, **43,** 260.
MOTT, N. F., 1948, *Proc. Phys. Soc.*, **60,** 391; 1953, *Phil. Mag.*, **44,** 742.
NATIONAL PHYSICAL LABORATORY, 1946, *J. Iron Steel Inst.*, **154,** (2), 83.
NOVICK, A. S., and MACHLIN, E. S., 1947, *J. Appl. Phys.*, **18,** 79.
OROWAN, E., 1934, *Z. Phys.*, **89,** 614.
POLANYI, M., 1934, *Z. Phys.*, **89,** 660.
READ, T. A., and TYNDALL, S. P. T., 1946, *J. Appl. Phys.*, **17,** 713.
SHUSHPANOV, L. I., 1937, *Metallurg*, **12,** 31.
TAYLOR, G. I., 1934, *Proc. Roy. Soc.* A, **145,** 362.
THOMPSON, D. O., 1952, *J. Appl. Phys.*, **23,** 1277.
WEINBERG, E. H., 1952, *J. Appl. Phys.*, **23,** 1277.

The Relation between Stress, Strain and Birefringence in Some High Polymers

By S. M. CRAWFORD

Imperial Chemical Industries Limited, Butterwick Research Laboratories,
Welwyn, Herts.

MS. received 16*th July* 1953

Abstract. Simultaneous measurements of the birefringence, stress and strain in crystalline and amorphous Terylene have been made at different temperatures. The results are compared with those obtained for polyethylene and Perspex where the birefringence is a single-valued function of strain. It is found that, for both the crystalline and amorphous forms, the birefringence is a function of both stress and strain and, for small strains, can be represented by an equation of the form $\Delta n \doteq A\sigma + B\epsilon$, where Δn is the birefringence, σ the stress, ϵ the strain and A and B are constants. The behaviour of these polymers is discussed in terms of this equation.

In an earlier paper (Crawford and Kolsky 1951) it was shown experimentally that the birefringence produced on stressing polyethylene is a single-valued function of strain. The birefringence of Perspex (a plasticized polymethyl methacrylate), which is amorphous, has also been observed to be a single-valued function of strain (Kolsky 1952).

Similar experiments have been carried out on Terylene (polyethylene terephthalate), which can be obtained in either an amorphous or a crystalline form, in order to compare its behaviour with that of these other materials. The amorphous form is only stable up to 80°c and is prepared by rapidly cooling the melt. If this form is heated above 80°c crystallization sets in (Reddish 1950). The crystalline form, melting point 264°c, is very brittle at ordinary temperatures, and it was found to be impossible to carry out experiments at the same temperatures with the two states of the material.

When the birefringence–strain graphs were plotted, it was found that the birefringence was not a single-valued function of the strain for either crystalline or amorphous Terylene. In the case of crystalline Terylene several specimens were taken through an hysteresis cycle at one temperature, and from the stress–birefringence and strain–birefringence graphs values for the birefringence and stress at constant strain were obtained. The birefringence was then plotted against the stress for each value of the strain and a family of curves was obtained from which it appeared that the birefringence was a single-valued function of stress and strain.

Filon and Jessop (1923) found that, for celluloid, a relation between the birefringence, stress and strain of the form $\Delta n = A\sigma + B\epsilon$ gave fairly good agreement with their experimental results. Δn is here the birefringence, σ the stress and ϵ the strain, A and B being constants. It was decided, therefore, to see if such an equation could be fitted to the results obtained with Terylene.

A value for B at any temperature was found from the residual birefringence and strain after unloading, and a value for A was then found from one point on the loading portion of the curve at low strains, using the value of B already obtained. At higher strains some difficulty was encountered in measuring the strain and birefringence simultaneously, on account of the creep. The value of the birefringence over the complete cycle of loading and unloading was then calculated from the equation given above, using the values of A and B derived as described and the values of the stress and strain from the stress–strain curve.

The birefringence–strain curves so calculated are shown in figs. 1 and 2 for amorphous and crystalline Terylene respectively, at two different temperatures. For comparison the experimental points have been inserted, and it is seen that good agreement is obtained except at the larger strains where it was difficult to obtain reliable experimental results.

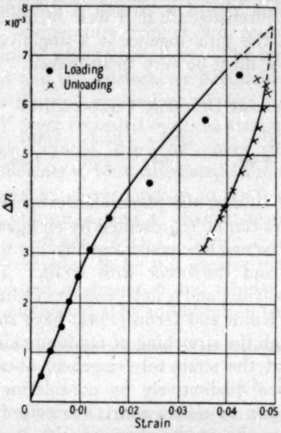


Fig. 1. Theoretical birefringence–strain curve calculated for amorphous Terylene at 39°C with experimental points shown for comparison.

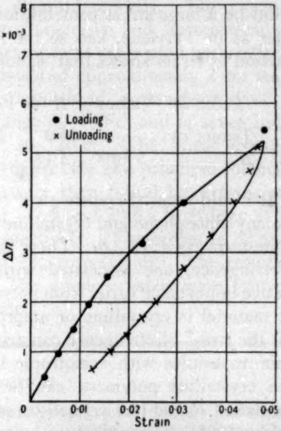


Fig. 2. Theoretical birefringence–strain curve calculated for crystalline Terylene at 81°C with experimental points shown for comparison.

Table 1. Values of A and B for Terylene at Different Temperatures

Form		Amorphous			Crystalline		
Temp (°C)		39·0	59·0	79·0	81·0	101·0	117·5
A (brewsters)	$\times 10^2$	1·04	1·58	98·97	1·17	1·15	1·13
B	$\times 10^{-2}$	8·80	8·08	4·35	5·26	5·00	4·45

Table 1 shows the values of A and B calculated for the two forms of the material at different temperatures. B is a dimensionless number since it is the ratio between birefringence and strain, whilst A has the dimensions of the reciprocal of stress, given in brewsters, where 1 brewster $= 10^{-13}$ dyn^{-1} cm^2. It can be seen that the values of A and B are of the same order in the two forms of the material except at the transition point where A, for the amorphous form, is some 100 times larger. This is probably associated with the fact that, at this temperature, Young's modulus is some 100 times less than at 39°C. With increasing temperature both A and B decrease in crystalline Terylene, but A appears to increase in the amorphous material while B decreases. Table 2 shows the values of B for polyethylene at different temperatures. It is seen to be of the same order as for Terylene but decreases more rapidly with increasing temperature as the 'softening point' at 110°c is approached.

The birefringence–strain graphs for polyethylene and Perspex will be represented by equations of the form $\Delta n = B\epsilon$ for small strains. The stresses required to produce strains in polyethylene of the same order as in Terylene were much smaller than those required for Terylene so that any stress contribution to the birefringence would be smaller than in Terylene for a similar value of A. It could still be detected, however, and the results show that A, for polyethylene, must be considerably smaller than for Terylene. In the case of Perspex there would be a large stress contribution to the birefringence if A were of the same order as in Terylene, and so the fact that the birefringence is a single-valued function of strain shows that A, for Perspex, must be very small indeed.

Table 2. Values of B for Polyethylene at Different Temperatures

Temp (°c)		19·0	31·0	50·5	70·2	80·2	90·7
B	$\times 10^{-2}$	5·12	4·96	4·40	3·64	2·07	1·86

It would seem therefore that, for small strains, the photo-elastic behaviour of many amorphous and crystalline polymers can be represented by an equation of the form $\Delta n = A\sigma + B\epsilon$. There appear to be two mechanisms giving rise to the birefringence, one associated with stress and the other with strain. Their relative importance varies from polymer to polymer and is independent of whether the material is crystalline or amorphous. Kuhn and Grün (1942) have shown that the stress-birefringence can arise through the stretching of randomly kinked chain molecules with anisotropic links, and the strain-birefringence, obtained with crystalline polymers, can be explained qualitatively by considering the orientation of rod-like crystallites embedded in an elastic matrix (Crawford and Kolsky 1951). It would seem, however, from the results obtained with Perspex and amorphous Terylene, that an orientation effect, as distinct from the stressing of bonds, may also occur in amorphous materials, whilst the birefringence produced in a crystalline material may depend on the stressing of bonds as well as orientation of the crystallites. This may well occur in crystalline Terylene because it contains more highly polar bonds than polyethylene.

ACKNOWLEDGMENT

The author would like to thank Dr. H. Kolsky for suggesting this work.

REFERENCES

CRAWFORD, S. M., and KOLSKY, H., 1951, *Proc. Phys. Soc.* B, **64**, 119.
FILON, L. N. G., and JESSOP, H. J., 1923, *Phil. Trans. Roy. Soc.* A, **223**, 89.
KOLSKY, H., 1952, *Trans. Soc. Glass Tech.*, **36**, 56.
KUHN, W., and GRÜN, F., 1942, *Kolloidzschr.*, **101**, 248.
REDDISH, W., 1950, *Trans. Faraday Soc.*, **46**, 459.

The Measurement of the Thermal Expansion of Single Crystals of Indium and Tin with a Photoelectric Recording Dilatometer *

By E. V. VERNON † and S. WEINTROUB

Physical Laboratory, The University, Southampton

MS. received 8th May 1953, *and in final form 22nd June* 1953

Abstract. An improved form of the interferometric dilatometer of Childs and Weintroub is described briefly and shown diagrammatically. The expansion and temperature fringes are recorded simultaneously and automatically by means of photomultipliers. The interferometer plates are coated with reflecting films and the dilatometer is suitable for crystals of approximately 2 cm length. The thermal expansions of two tin single crystals of orientations $\psi = 8$ and $65{\cdot}5°$ were recorded over the temperature range 30–180°c, and of seven indium single crystals, $\psi = 7$–$82°$, over the range 30–120°c. The linear expansion coefficients α_ψ were calculated from the records and the two principal coefficients $\alpha_\parallel$ and $\alpha_\perp$ corresponding to $\psi = 0$ and 90° were then found by extrapolation. For indium, $\alpha_\parallel$ and $\alpha_\perp$ change steadily from $-7{\cdot}5$ and $50{\cdot}0 \times 10^{-6}$/deg c at 30°c to $-29{\cdot}0$ and $72{\cdot}5 \times 10^{-6}$/deg c at 120°c respectively. The errors in $\alpha_\parallel$ and $\alpha_\perp$ are estimated to be not greater than 5 and 1% respectively. The results are exhibited in graphical and tabular form. The single crystals were grown by the method of slow solidification from the melt.

§ 1. INTRODUCTION

GENERAL information on the anisotropic properties of metals has recently been summarized by Boas and Mackenzie (1950) and on the anisotropy of the thermal expansion, in particular, by Childs (1949, 1953). These publications show the need for additional accurate experimental data on the variation with temperature of the principal thermal expansion coefficients of metals. Recent developments in the techniques of the production of pure metal single crystals and of the determination of their orientations (Goss and Weintroub 1952) have facilitated this experimental investigation. The application of the Fizeau optical interference method to the measurement of the thermal expansion of anisotropic metals (Childs and Weintroub 1950) has also assisted. However, the apparatus used by Childs and Weintroub had the disadvantage that continuous and tedious visual observation of fringe movements was required. Photographic recording has been used by several investigators, e.g. Saunders (1945), but automatic photoelectric recording is considered by the present authors to be better.

In this paper a brief description is given of an improved form of apparatus, in which the essential and best features of previous optical interference dilatometers have been retained, but in which the novel* feature of automatic and continuous photoelectric recording of both the expansion and temperature has been incorporated. With this apparatus the previous measurements of Childs and Weintroub on the thermal expansion of tin have been augmented by an 8° and a 65·5° crystal, and detailed measurements over the temperature range 30–120°c have been made on seven crystals of indium of orientations 7–82°. The results for indium differ considerably from the very few previously published values. The two principal linear thermal expansion coefficients, $\alpha_{\parallel}$ and $\alpha_{\perp}$, corresponding to the directions parallel and perpendicular to the *c*-axis of indium, were determined by extrapolation, and found to change steadily from $-7{\cdot}5$ to $-29{\cdot}0 \times 10^{-6}$/deg c, and from 50·0 to $72{\cdot}5 \times 10^{-6}$/deg c respectively over the range 30–120°c. The negative value of $\alpha_{\parallel}$ has not previously been reported. The magnitude however is in agreement with former values.

---

* The work described in this paper forms the substance of a thesis for which the degree of Ph.D. of the University of London was awarded in March 1953 to E. V. Vernon.

† Now at British Callender's Cables Ltd., London.

§ 2. THE APPARATUS

The general layout of the photoelectric recording interference dilatometer is shown in fig. 1. The instrument was designed and constructed to record simultaneously the change of length and temperature of metal single crystals about 2 cm long, when the temperature was varied steadily between room temperature and the melting point of the metal. The principle is that of Fizeau (1864) and the form of the apparatus is similar to that of Saunders (1945) except that photoelectric recording is used in place of photographic recording. The photoelectric is, in our opinion, better than the photographic method, both in accuracy and in ease of operation.

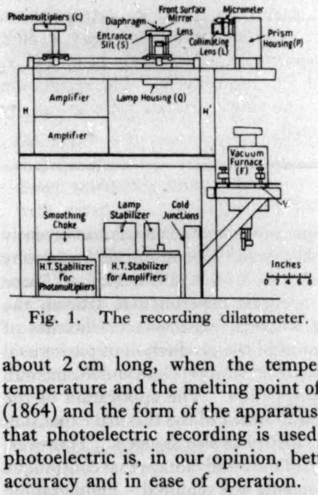

Photomultipliers (C)
Diaphragm
Front Surface Mirror
Micrometer
Entrance Slit (S)
Lens
Collimating Lens (L)
Prism Housing (P)
Amplifier
Lamp Housing (Q)
H
H'
Amplifier
Vacuum Furnace (F)
Y
Smoothing Choke
Lamp Stabilizer
Cold Junctions
H.T. Stabilizer for Photomultipliers
H.T. Stabilizer for Amplifiers
Inches
0 2 4 6 8

Fig. 1. The recording dilatometer.

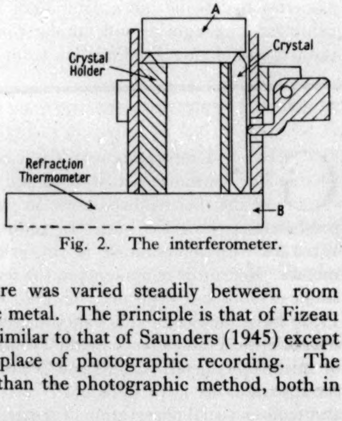

A
Crystal
Crystal Holder
Refraction Thermometer
B

Fig. 2. The interferometer.

The main components of the dilatometer were the vacuum furnace F, containing the interferometer (fig. 2) with its three crystal spacers and refraction thermometer, the constant deviation prism and mounting P, the collimator slit S and light source in the housing Q, and the two RCA 931A photomultipliers C and their associated circuits. These components were attached to the rigid steel framework HH'. The optical arrangement is shown in fig. 3.

---

* During the course of the work described in this paper a report of an apparatus also using the photoelectric method was published independently by Work (1951).

The Interferometer

The interferometer consisted of two horizontal optically flat glass discs A and B about 2 cm apart, separated by three spacers of very nearly equal length. These spacers were portions cut from the same metal single crystal rod. They were maintained accurately vertical, but without undue constraint, by a specially designed crystal holder. The upper flat A, which rested on the tops of the spacers, was 15 mm in diameter and 5 mm thick and was slightly prismatic with a wedge angle of 15 minutes of arc. The lower flat B was 38 mm in diameter and 5 mm thick, and its underneath surface was divided into two portions. The one half, over which the crystal holder rested, was ground to stop reflection, and the other half was made prismatic with a wedge angle of magnitude such that when this portion of B was illuminated from above by normally incident green light five fringes were visible. This half constituted the refraction thermometer (Luckeish *et al.* 1922).

The lower surface of A and the upper surface of the refraction thermometer were coated with 30% reflecting films of zinc sulphide. The lower surface and the portion of the upper surface of B below A were coated with 60% reflecting films of aluminium. The evaporation method used for the deposition of the zinc sulphide has been briefly described by Vernon (1953 a). The reasons for the choice of such films and the improvement in the contrast of the fringes produced by them are discussed by Vernon (1953 b).

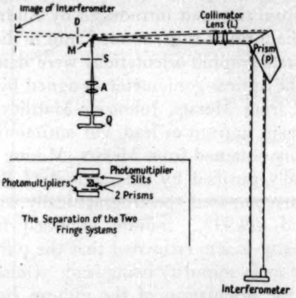


Fig. 3. Ray diagram of the optical system.

The Vacuum Furnace

The furnace was similar to that used by Childs and Weint oub (1950). The electrical energy supplied to the furnace was controlled so that the temperature increased or decreased approximately linearly with time. The control unit was based on that of Adcock (1935). The rate of temperature change could be varied within the limits 2–8°c h^{-1}.

The Source and Record

A mercury capillary 'H' form discharge lamp, operated from a stabilized 50 c/s supply, was used as the source of illumination, and the prism P and optical system were adjusted so that accurately collimated monochromatic light from the blue 4358 Å line of the source was incident normally on the interferometer placed within the furnace. Some five to ten interference fringes were formed by the light multiply reflected between the lower surface of A and the upper surface of B. The refraction thermometer was calibrated by means of a chromel–alumel thermocouple which was located in the crystal holder.

The images of the two fringe systems were separated by means of the two prisms (fig. 3, inset), and brought to a focus one on each of the photomultiplier slits, which were about 1 mm long, and which were orientated parallel to the fringes. An RCA 931A photomultiplier was located behind each slit and the resulting signal corresponding to the intensity of the fringe illuminating the slit was fed to two cascaded stages of a stable 'twin-T' selective amplifier (Sturtevant 1947) tuned to 100 c/s. The variations in output corresponding to the passage of the two sets of fringes across the slits were recorded together on an Evershed and Vignoles Duplex recorder.

§ 3. THE SINGLE CRYSTAL SPACERS

The two tin crystals, of orientations 8° and 65½°, and the seven indium crystals, of orientations 7, 18, 32½, 41, 52½, 72½ and 82°, were grown by the method of slow solidification from the melt introduced by Andrade and Roscoe (1937). The crystals were in the form of rods, about 3 mm in diameter and up to 15 cm in length. Their crystallographic orientations were determined by the etch-pit reflection method. The optical goniometer designed by Goss (1953) was used. The tin was obtained from Messrs. Johnson, Matthey & Co. Ltd., and was 99·997% pure, with main impurities lead and antimony. Commercial indium of 99·9 to 99·95% purity obtained from Messrs. Mining and Chemical Products Ltd., was electrolytically purified by the method of Baxter and Alter (1933). The purified metal was compared spectrographically with a sample of spectroscopically standardized (99·99%) indium obtained from Messrs. Johnson, Matthey & Co. Ltd., and it was estimated that the purified metal was not less than 99·98% pure, the main impurity being lead. Details of the growth and of the determination of the orientation of the indium single crystals have been published elsewhere (Goss and Vernon 1952).

Conversion into Spacers

Three spacers, equal in length to about 1 μ, were produced from each crystal. The method used was devised to avoid mechanical strain, and it is considered to be an improvement in this respect on that previously used by Childs (1950). A 3 mm × 3 mm groove was machined down the length of a 8 in. × ½ in. × ½ in. steel bar, and saw cuts were made at intervals of 22 mm (for tin) and 24 mm (for indium) across the groove. The metal single crystal was fixed into the groove by distrene cement and cut into the required lengths, 22 or 24 mm, with a fine saw. The spacers were then detached from the steel bar by dissolving the cement away with toluene. Traces of distrene remaining on the crystal spacers were removed by immersing the spacers for a short time in hydrochloric acid.

The recrystallization formed at the ends of the spacers where the sawing had taken place was removed by electropolishing some 2 mm off the ends of the spacers by immersion in suitable solutions as recommended by Jacquet (1939) for tin and Carapella and Peretti (1949) for indium. The electropolishing not only removed the recrystallized material but also produced domed ends, which were very suitable for the three-point support of the glass disc A. In addition it facilitated the accurate adjustment of the spacers to equal lengths without making use of mechanical working. The spacers, when finally assembled in the interferometer, were annealed at a temperature of approximately 200°c for tin and 130°c for indium.

§ 4. The Thermal Expansion of Tin and of Indium

The thermal expansions of the two tin crystals ($\psi = 8°$ and $65{\cdot}5°$) were measured over the range of temperature 30–180°c. The variations of α_ψ, the linear expansion coefficient, with temperature is shown in fig. 4, together with the corresponding values, for the crystals of orientations $\psi = 86\frac{1}{2}, 79\frac{1}{2}, 44\frac{1}{2}, 43\frac{1}{2}, 31\frac{1}{2}$ and 30°, obtained by Childs and Weintroub (1950).

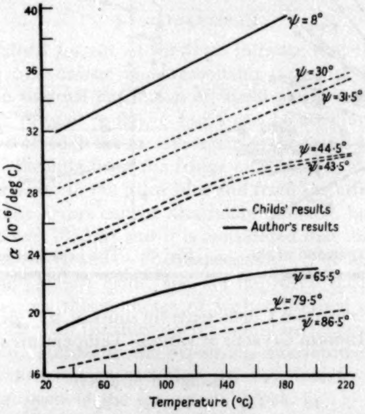


Fig. 4. The variation of expansion coefficient with temperature for tin.

Several runs were made on the seven indium crystals of orientations $\psi = 7, 18, 32\frac{1}{2}, 41, 52\frac{1}{2}, 72\frac{1}{2}$ and 82°. Each run consisted of heating the spacers from room temperature to 120°c and then cooling to room temperature again. The cycle was carried out at a rate of about $4\frac{1}{2}$°c per hour. In those cases where it was found that because of slight plastic flow at the tops of the spacers the expansions for heating and cooling were different, the run was repeated until the results for heating and cooling agreed to within 2%. For the crystal of orientation $\psi = 52\frac{1}{2}°$ better agreement than to 8% could not be obtained; this is attributed to either slip within the crystal or to extreme softness. In fig. 5 the crystal expansion fringes for the various crystals are plotted against the refraction thermometer fringes, and the expansion coefficients were calculated by taking the slopes, at 10°c intervals, of these curves and of the calibration curve of refraction thermometer fringes against temperature. The calibration curve was approximately linear. Values for $\alpha_\parallel$ and $\alpha_\perp$ were deduced by extrapolation, and these are included in the values of α_ψ listed in table 1. The variation of $\alpha_{\parallel}$ and $\alpha_{\perp}$ with temperature is shown in fig. 6.

The negative expansion coefficient for the crystals of low orientation was confirmed by growing a long (20 cm) single crystal rod of 7° orientation and measuring, by means of a comparator, the changes in the length between two fiduciary marks on the rod, when the crystal was heated from room temperature to approximately 100°c and then cooled again. The mean expansion coefficient obtained was in good agreement with the data previously obtained.

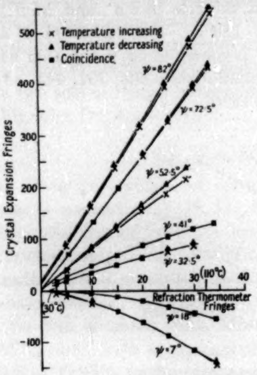


Fig. 5. The expansion of the indium single crystals.

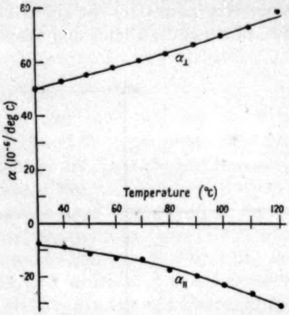


Fig. 6. The principal expansion coefficients of indium as functions of temperature.

Table 1. The Expansion Coefficients (in units of 10^{-6}/deg c) of the seven Indium Crystals at various Temperatures

Temp. (°c)	Orientation (degrees)								
	0	7	18	32½	41	52½	72½	82	90
30	− 7·5	− 7·5	− 1·4	12·5	16·2	27·2	44·4	49·7	50·0
40	− 9·0	− 9·1	− 2·7	12·2	15·7	28·4	47·0	51·8	52·5
50	−11·0	−10·9	− 3·9	11·8	15·3	29·5	48·8	53·2	54·5
60	−13·0	−12·8	− 5·1	11·5	14·9	30·5	50·6	55·2	57·0
70	−14·0	−14·9	− 6·4	11·0	14·5	31·3	52·4	57·3	59·0
80	−17·0	−17·2	− 7·7	10·4	14·0	32·1	54·3	59·7	61·0
90	−19·0	−19·6	− 9·2	9·8	13·6	32·8	55·9	62·2	62·5
100	−20·5	−22·2	−10·9	9·3	13·2	33·4	57·2	64·7	65·0
110	−24·5	−25·2	−12·9	8·5	12·8	34·0	58·4	68·0	68·0
120	−29·0	−28·5	−13·4	—	12·4	34·6	—	71·6	72·5

The values of 0° ($\alpha_{\parallel}$) and 90° ($\alpha_{\perp}$) were obtained by extrapolation.

In table 2 the values of $\alpha_{\parallel}$ and $\alpha_{\perp}$ obtained in this investigation are compared with those determined by other investigators. In neither of the two x-ray determinations was the negative sign of $\alpha_{\parallel}$ mentioned, but the magnitudes, as can be seen from the table, are of similar order to those of the present work.

The only other reliable expansion measurement on indium is that of Hidnert and Blair (1943) who determined the expansion of polycrystalline indium for the range 20–100°c. Their values are included in table 2, together with values of $\alpha_{\text{polycrystal}} = \frac{1}{3}\alpha_{\parallel} + \frac{2}{3}\alpha_{\perp}$ calculated from the results of Shinoda (1933), Frevel and Ott (1935), the present investigation, and, for completeness, Fizeau (1869).

Table 2. Comparison between Values of α ($\times 10^{-6}$/deg c) for Indium found by various Observers and the Authors

Observer	Metal purity %	Temp. (°c)	$\alpha_{\parallel}$	$\alpha_{\perp}$	α polycrystal
Fizeau (1869)		40			41·7
		0–100			45·9
Frevel and Ott (1935)	99·9	−25–141	13·1	56·1	41·8
Hidnert and Blair (1943)	99·9	20–100			30·5
Shinoda (1933)		23–87	11·7	45·0	33·9
Authors	99·98	30	−7·5	50·0	30·8
	99·98	90	−19·0	62·5	35·4

§ 5. DISCUSSION

The results obtained for tin and indium indicate that the dilatometer in its present form is a convenient and reasonably accurate instrument for the measurement of the thermal expansion of metal single crystals. The random experimental scatter present in figs. 5 and 6 can be attributed to the combined effect of the errors involved in the measurement of the traces and in the calculation of the expansion coefficients from the fringe curves, and, also, the errors which arise from the uncertainty in the value of ψ and from imperfections in the crystal spacers. Not all these errors can be accurately assessed, but a careful estimate has been made (Vernon 1953 b) and it is considered that the errors in $\alpha_{\parallel}$ and $\alpha_{\perp}$ are not greater than ±5% and ±1% respectively. The temperature measurement introduces the greatest error, and for improved accuracy experience has shown that a refraction thermometer of higher thermal conductivity and in better thermal contact with the dilatometer would be necessary. It is suggested that an interferometer similar to the expansion interferometer be used in place of the refraction thermometer and that the two interferometers be in good thermal contact, e.g. parts of the same metal block.

The ratio $c/a = 1{\cdot}076$ for indium, and thus its structure, although tetragonal, is almost cubic. It is therefore difficult to account for the observed marked anisotropy in the thermal expansion. Very little is known about the anisotropy of the other physical properties of indium. Verhaeghe, Le Compte and Vandermeersssche (1951) found the ratio, $\chi_{\parallel}/\chi_{\perp}$, of the diamagnetic susceptibilities at room temperature to be 0·44, but some measurements made recently in this laboratory (Goss, unpublished) on the electrical resistivity of indium did not indicate any marked anisotropy. The elastic constants, as Grüneisen and Goens (1924) have pointed out, are closely related to the thermal expansions, and it would therefore be extremely valuable to know the values of these constants. Unfortunately, however, the only data available (Ludwick 1950) are incomplete.

ACKNOWLEDGMENTS

The authors are greatly indebted to the University authorities, and to Professor A. M. Taylor, in whose department the work was carried out, for the facilities provided. One of us (E. V. V.) also gratefully acknowledges the receipt of a maintenance grant from the University.

References

Adcock, F., 1935, *J. Sci. Instrum.*, **12,** 285.

Andrade, E. N. da C., and Roscoe, R., 1937, *Proc. Phys. Soc.*, **49,** 152.

Baxter, G. P., and Alter, C. M., 1933, *J. Amer. Chem. Soc.*, **55,** 1943.

Boas, W., and Mackenzie, J. K., 1950, *Progress in Metal Physics*, Vol. II (ed. Chalmers), (London : Butterworths), p. 90.

Carapella, S. C., and Peretti, E. A., 1949, *Metal Progr.*, **56,** 666.

Childs, B. G., 1949, *Dissertation*, University of London ; 1950, *J. Sci. Instrum.*, **27,** 102; 1953, *Rev. Mod. Phys.*, **25,** in the press.

Childs, B. G., and Weintroub, S., 1950, *Proc. Phys. Soc.* B, **63,** 267.

Fizeau, A. H. L., 1864, *Ann. Chim. Phys.*, 4th series, **2,** 143 ; 1869, *C. R. Acad. Sci., Paris* **68,** 1125.

Frevel, L. K., and Ott, E., 1935, *J. Amer. Chem. Soc.*, **57,** 228.

Goss, A. J., 1953, *J. Sci. Instrum.*, **30,** 283.

Goss, A. J., and Vernon, E. V., 1952, *Proc. Phys. Soc.* B, **65,** 905.

Goss, A. J., and Weintroub, S., 1952, *Proc. Phys. Soc.* B, **65,** 561.

Grüneisen, E., and Goens, E., 1924, *Z. Phys.*, **29,** 141.

Hidnert, P., and Blair, M. G., 1943, *Bur. Stand. J. Res., Wash.*, **30,** 427.

Jacquet, P. A., 1939, *Publication International Tin Research Development Council*, No. 90.

Luckeish, M., Holladay, L. L., and Sinden, R. H., 1922, *J. Franklin Inst.*, **194,** 251.

Ludwick, M. T., 1950, *Indium* (New York : Indium Corporation of America).

Saunders, J. B., 1945, *Bur. Stand. J. Res., Wash.*, **35,** 157.

Shinoda, G., 1933, *Kyoto Coll. Sci. Mem.* A, **16,** 193.

Sturtevant, J., 1947, *Rev. Sci. Instrum.*, **18,** 124.

Verhaeghe, J., Le Compte, G., and Vandermeerssche, G., 1951, *Meded. K. Vlaamse Acad. Wetensch. Belg.*, **13,** No. 8.

Vernon, E. V., 1953 a, *J. Sci. Instrum.*, **30,** 292; 1953 b, *Thesis*, University of London.

Work, R. N., 1951, *Bur. Stand. J. Res., Wash.*, **47,** 80.

RESEARCH NOTES

The Study of Lightning Streamers with 50 cm Radar

By F. J. HEWITT

Telecommunications Research Laboratory, South African Council for Scientific and Industrial Research, Johannesburg, South Africa

Communicated by B. F. J. Schonland; MS. received 22nd June 1953

§1. INTRODUCTION

VARIOUS workers (Browne 1951, Miles 1952) using wavelengths of 3 and 10 cm have shown that radar reflections can be obtained from the channels of lightning discharges; these observations have been confirmed by the writer. The presence of echoes from hydrometeors in the thundercloud, however, makes such short wavelengths unsuitable for the study of the propagation of lightning streamers within the cloud. As longer wavelengths appeared desirable for other reasons, equipment operating on a wavelength of 50 cm has been built and provided with a means of photographing the echoes from successive individual pulses. It was found that the recording system need not be run continuously but could be set in operation by the noise radiation, at the operating frequency of the radar, associated with each of the separate strokes of a discharge.

§2. EXPERIMENTAL OBSERVATIONS

The equipment proved capable of recording echoes from lightning up to a range of about 30 km and was not seriously affected by echoes from hydrometeors. The majority of the observations were made at ranges of about 8 km. This enabled the slant ranges measured by the equipment to provide some discrimination in height which would otherwise have been lacking, as the beam width of the equipment was too large (20°) for measurement of angle of elevation.

Early on in the work it was observed that when a lightning discharge occurred the noise radiated from certain stages of the discharge was clearly detectable on the radar set. A trigger system was therefore built to operate on this radiated noise. This trigger system made use of the radar aerial and receiver themselves so as to obtain approximately the same directional characteristics as the radar, and so to reduce undesirable triggering to a minimum.

The records show that noise radiation always precedes the appearance of radar echoes, and it is therefore assumed, as may be expected, that it is associated with the start of each stroke. The power flux of this radiation was observed to be at least $10^{-12}\,\mathrm{w\,m^{-2}}$ in a bandpass of 1 Mc/s centred on 600 Mc/s, at ranges of 8 to 16 km from typical lightning strokes. The duration of the radiation at this level was observed to lie between 100 μsec and 3 msec, and it appears that it may take two different forms. The first and most common type results in a general increase in receiver noise level, frequently rising to a maximum in under 10 μsec, while the second type appears to consist of a long series of short individual spikes at intervals of a few microseconds.

The recording system consisted of a drum camera running continuously and capable of recording echoes from successive transmitter pulses at intervals of one millisecond, for a period of 0·5 sec after the operation of the trigger. Range–amplitude presentation was used.

The echoes from discharges which took place entirely within the cloud were complicated, and in general extended over several kilometres in range. Various components of these echoes frequently persisted for more than 100 msec, but no obvious trend could be seen except in those cases where strokes to ground appeared to follow the cloud discharge.

In the case of discharges to ground a distinctive series of echoes was observed to follow each successive stroke. On a number of occasions the echoes could be resolved into two main components separated in slant range by about a kilometre. One of these components, usually the short-range component, appeared at its greatest strength immediately after the stroke and faded to a small fraction of its initial intensity in about 10 msec. The other component was observed to build up over the 10–50 msec and then gradually diminish, to disappear immediately before the commencement of the next stroke.

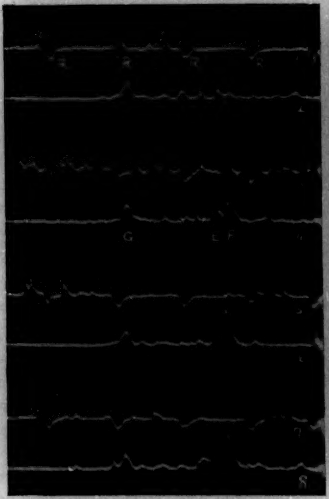


Fig. 1. Commencement of stroke.

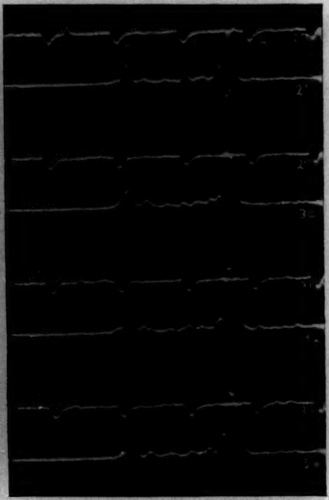


Fig. 2. 12 milliseconds later.

An example of this effect is shown in figs. 1 and 2, which refer to the same stroke at its start and 12 msec later, respectively.

The duration of each time-base on the record was just over 40 μsec and the time interval between the start of successive time-bases was 500 μsec. Radar echoes appear only on evenly numbered time-bases and the small echo G is an echo from some prominent ground feature. This echo G appears on all evenly numbered time-bases and is not associated with the lightning flash.

The calibration marks R appear on odd numbered time-bases only. They are at intervals of 10 μsec, corresponding to radar range intervals of approximately 1·5 km, the range increasing from left to right. The ground echo G is at a range of just over 6 kilometres. The calibration marks do not appear on the evenly numbered time-bases to avoid confusion with possible radar echoes.

Noise picked up by the receiver aerial may appear on any of the time-bases.

In fig. 1 a burst of noise radiated by the lightning stroke is seen on time-base 3. On time-base 4, 500 μsec later, a composite echo, with two main components E and F, has appeared. This composite echo is not present on time-base 2 to any appreciable extent.

As time elapses this composite echo changes in type until 13 msec later, on time-base 30, the component E has practically disappeared while the component F has increased considerably, reaching a maximum value at about this time. The component F was then observed to fade gradually until it disappeared about 50 msec after its first appearance on time-base 4. The next stroke in the flash occurred 56 msec after time-base 4.

It is tentatively suggested that the echo component E referred to above is associated with the ionization persisting in the main channel after each stroke and that the component F may be associated with the ionization resulting from the junction streamer process that takes place between successive strokes as described by Malan and Schonland (1951).

On certain records it was observed that the range of the echo tentatively associated with the junction streamer process increased slightly with time in steps from one stroke to the next. On the assumption of a vertical discharge path the mean vertical velocity of the region concerned ranged between 4 and 10 km sec^{-1}. On other records no such increase in range was observed. These studies are being continued.

§3. CONCLUSIONS

The fact that radar echoes can be observed during the whole of the period between successive strokes in a lightning flash, and that some components of these echoes appear to be associated with the processes leading up to the next stroke, suggests that radar may provide a useful means of studying these processes provided sufficient resolution in range and angle of elevation can be obtained.

ACKNOWLEDGMENTS

The author wishes to express his thanks to the South African Council for Scientific and Industrial Research for permission to publish this Note and to Dr. B. F. J. Schonland, who suggested this research and provided accommodation at the Bernard Price Institute, for his encouragement and advice.

REFERENCES

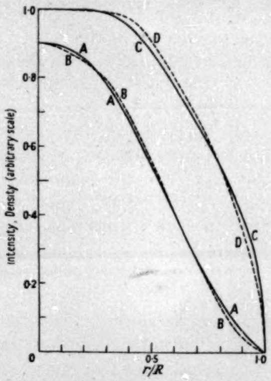
BROWNE, I. C., 1951, *Nature, Lond.*, **167**, 438.
MALAN, D. J., and SCHONLAND, B. F. J., 1951, *Proc. Roy. Soc.* A, **206**, 145.
MILES, V. G., 1952, *Nature, Lond.*, **170**, 366.

LETTERS TO THE EDITOR

Spectroscopic Study of Caesium Discharges in a Magnetic Field

In a previous paper (Davies 1953, to be referred to as I) the electron density and temperature in a caesium discharge with a longitudinal magnetic field were investigated by means of the electron–ion recombination radiation. This note reports the results of a similar investigation using a positive column of larger diameter (1·8 cm). The discharge current and vapour pressure were 2·5 A and 0·087 mm Hg respectively in all cases.

As before, the electron velocity distribution was found to be Maxwellian. A magnetic field of 1200 gauss produced no change in electron temperature for either direct current or radio-frequency excitation (6 Mc/s); a change of 2% would have been detectable. This result is contrary to the predictions of diffusion theory (Tonks 1939) as pointed out in I. For the direct current case no change in electron temperature or axial density was observed when the current direction was reversed, either with or without a magnetic field. In I a small change was observed, which is probably accounted for by the much greater current density used in the tube of smaller diameter; this effect may also be related to the increase in electron temperature produced by the magnetic field which was observed in I in radio-frequency discharges, but was not observed here.



Intensity distribution across the spectral image of the positive column (curves A, B), and derived distribution of electron density (curves C, D). Curves A, C are for d.c. and radio-frequency discharges with zero magnetic field; curves B, D are for the radio-frequency discharge in a longitudinal field of 1200 gauss.

For the case in which the electrons have a Maxwellian distribution and the electron temperature is independent of radius, the intensity of the recombination continuum is proportional to the square of the electron density. The radial variation of electron density can thus be found by forming an image of the positive column on the spectrograph slit and measuring the variation of intensity in the spectral image at a wavelength in the continuum. These results were interpreted using an integral equation due to Hörmann (1935), assuming the self-absorption to be negligible. The magnetic field (1200 gauss) increased the axial electron density by 10% for d.c. and 4% for radio-frequency current; the accuracy of measurement was to within about 5%.

In the absence of a magnetic field both d.c. and radio-frequency discharges had the same radial density distribution (curve C). Application of the field to the radio-frequency discharge caused a 'broadening' of the distribution (curve D). A similar broadening was observed when the discharge was run at very low current (500 μA d.c.) in a tube 5 mm in diameter; in the absence of the field the positive column took the form of a luminous 'cord' of 0·85 mm diameter, which broadened to 1·35 mm when a field of 1200 gauss was applied, without any visual change in intensity. The radial density distribution in a d.c. discharge in the presence of a magnetic field was not investigated.

The electron density distribution is seen to differ from the Schottky (1924) diffusion theory of the positive column (zero-order Bessel function). This is not surprising, since volume recombination and electron scattering by positive ions undoubtedly affect the diffusion relations in this case.

The experiments described here were carried out at the Clarendon Laboratory during the tenure of a Rhodes Scholarship. Thanks are due to Drs. H. G. Kuhn and P. C. Thonemann for help with the manuscipt.

Division of Radiophysics, Commonwealth Scientific
and Industrial Research Organisation,
University Grounds, Sydney, Australia.
L. W. Davies.

27th February 1953, and in final form 5th August 1953.

Davies, L. W., 1953, *Proc. Phys. Soc.* B, **66,** 33.
Hörmann, H., 1935, *Z. Phys.*, **81,** 571.
Schottky, W., 1924, *Phys. Z.*, **25,** 342, 635.
Tonks, L., 1939, *Phys. Rev.*, **56,** 360.

Surface Recombination in Germanium

An investigation has been made of the variation of the surface recombination velocity of injected minority carriers in germanium, over a limited temperature range. The method used was that of Suhl and Shockley (1949), and is indicated in fig. 1 (*a*). A longitudinal electric field E_1 is maintained in a long thin rectangular block of germanium. Minority carriers are injected into the block by an emitter point E, and swept towards a collector point C. A variable magnetic field H is applied in a direction perpendicular to the plane on which the emitter point is situated. The injected carriers are thereby deflected towards or away from the collector point, depending on the sign of H. The number of such carriers actually arriving at the collector depends on the characteristics of the surface. The surface recombination velocity s, which is a measure of the surface decay rate, can be evaluated in terms of the relative concentration of minority carriers near the collector point, as shown by Suhl and Shockley. The most convenient way of obtaining the value of this concentration is to measure the point conductance of C about zero current for an applied voltage less than kT/e in each direction. This method (Bardeen 1950) eliminates the effect of changes in the potential of C due to magneto-resistance and alterations in floating potential.

Figure 1 (*b*) shows room-temperature results of such measurements for an n-type and a p-type specimen. G_0 is the point conductance in the absence of injection and magnetic field, G_1 with injection but no magnetic field and G with both electric and magnetic fields. The curve for the n-type specimen is very similar to that reported by Suhl and Shockley and the results confirm that the surface recombination properties of p-type material are closely analogous, subject to the usual reversal of signs. Tests were made to confirm that the measurements were not disturbed by carrier injection at the end contacts of the germanium filament, nor by heating due to the currents flowing. The emitter and collector were also used to determine the sweeping field. This was done approximately by switching off the emitter current while maintaining the sweeping current at its original value, and then using the emitter and collector as voltage probes.

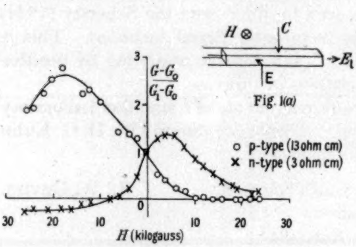


Fig. 1 (*b*).

Fig. 1 (*a*). Arrangement of electrodes and fields for measurement of the Suhl effect. (*b*). Relative conductance of collector point as a function of magnetic field.

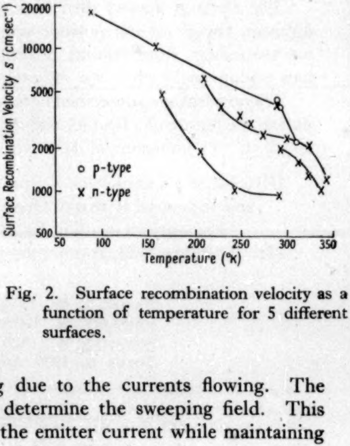


Fig. 2. Surface recombination velocity as a function of temperature for 5 different surfaces.

Figure 2 shows the corresponding values of the surface recombination velocity as a function of temperature. (Readings below room temperature were taken in a vacuum incorporating a liquid air trap.) These results were obtained on moderately well-etched surfaces. It was found that the absolute values are very sensitive to changes of etching procedure. The measurements by this method become much more difficult for surface recombination velocities outside the present range of values. The diminution of *s* with increasing temperature is compatible with the assumption that carriers are first trapped in the surface and that the probability of recombination increases with the time spent in the traps (cf. Brattain and Bardeen 1953, Lawrance 1953). The calculations require a knowledge of the carrier mobility as a function of temperature and the results obtained by Lawrance were used.

The author wishes to thank Professor R. W. Ditchburn for placing research facilities at his disposal, Mr. T. R. Scott and Mr. R. W. Douglas for the supply of specimens, and the Royal Naval Scientific Service for their support.

Physics Research Laboratory,
University of Reading.
29th July 1953.

W. N. Reynolds.

Bardeen, J., 1950, *Bell Syst. Tech. J.*, **29**, 469.
Brattain, W. H., and Bardeen, J., 1953, *Bell Syst. Tech. J.*, **32**, 1.
Lawrance, R., 1953, *Phys. Rev.*, **89**, 1295.
Suhl, H., and Shockley, W., 1949, *Phys. Rev.*, **75**, 1617.

REVIEWS OF BOOKS

Superconductivity, 2nd Edn., by D. Shoenberg. Pp. x + 256. Cambridge Monograph on Physics. (Cambridge: University Press, 1952.) 30*s*.

Although the phenomenon of superconductivity has until now maintained its position as an unsolved mystery of science, so much new experimental and theoretical evidence has accumulated since the publication in 1938 of the first edition of Dr. Shoenberg's monograph that in its second edition the book has had to be greatly enlarged and revised. An introductory chapter deals with the discovery of the effect (it was a happy thought to include the delightful portrait sketch of the discoverer) and rapidly surveys the outlines of the subject. The account which follows of the magnetic and thermal properties of macroscopic superconductors reproduces the main topics of the first edition, but has been rearranged and expanded to include recent results on thermal conductivity and thermoelectric effects. The next chapter gives an account of the recent investigations, experimental and theoretical, which have elucidated the structure of the intermediate state; of particular interest here are the remarkable Russian experiments which have demonstrated directly that the metal in the intermediate state is broken up into a complicated mixture of superconducting and normal regions. A further chapter describes the various methods which have been used to study the depth of penetration of a magnetic field into a superconductor and which have provided much important information, not only on the absolute value of the penetration depth, but also on its variation with temperature, crystal orientation and magnetic field strength. The last chapter, which is of great interest, discusses the extent to which the phenomenological theories of superconductivity, the thermodynamic theory of Gorter and Casimir and the electrodynamic theory of F. and H. London have been successful in correlating the experimental facts, and emphasizes that recent experiments, particularly the important work of Pippard on the behaviour at high frequencies, have revealed phenomena which seem to lie beyond the grasp of these theories. A brief sketch follows of the various attempts at a fundamental theory of superconductivity which have lately been appearing in regular succession; although it seems fairly certain now that superconductivity is in some way connected with the interaction between the electrons and the lattice vibrations in a metal, no convincing demonstration has yet been given of the way in which this interaction can lead to the characteristic properties of a superconductor. An appendix gives an extensive list of numerical data on superconducting transition temperatures and critical fields.

It is a pleasure to read Dr. Shoenberg's clear and critical account of the present state of our knowledge of what is surely among the most remarkable of all physical phenomena. The book can be strongly recommended to the general reader interested in the subject, and will be indispensable to the specialist. It is evident that, between the first and second editions of the book, the plot of the story has greatly thickened; will the third edition be able to announce that it has at last been unravelled?

E. H. SONDHEIMER.

Vacuum Technique, by ARNOLD L. REIMANN. Pp. ix+449. (London: Chapman and Hall, 1952.) £2 10*s*. 0*d*.

This book may be classified as filling the gap in the range of modern textbooks on vacuum practice in English which exists between the smaller monographs on the subject and the comprehensive work *Scientific Foundations of Vacuum Technique* by Dushman. As such, it is a very useful addition to the literature, is well produced and excellently illustrated, and can be readily understood without too great a demand on the reader's knowledge of mathematics and physics.

In the text are included comprehensive accounts of backing, molecular and diffusion pumps, vacuum gauges, speed measurements, pump systems, absorption of residual gases, glass and glass–metal seal making. Especially praiseworthy features are the detailed drawings of apparatus which should be of great use to the glass-blower or technician faced with the task of fabricating the research worker's vacuum apparatus, and also the many useful practical hints given concerning such matters as glass working, seal making, pump oils, waxed and cement joints, to mention but a few amongst many.

The only major criticism of the text is that it includes little—apart from fundamental considerations—which is of direct interest to the industrial users of vacuum apparatus other than those engaged in work on electron tubes. Doubtless this is because the author wished to write chiefly of those vacuum matters of which he had had first-hand experience. Probably this was a wise restriction of the extent of the text, for the reader is soon convinced that here is an excellent account of vacuum technique as practised in the laboratory, and that the advice and instructions given are the result of many years of research experience.

It is strongly recommended as an account of vacuum technique which, whilst emphasizing practice, yet includes a wisely selected account of underlying theory. It should prove valuable to the physicist, the technician, and the student beginning research; it will undoubtedly be consulted on many occasions by the workers in University and technical college laboratories where vacuum apparatus must be frequently designed and built to avoid the expense of purchasing manufacturers' equipment.

J. Y.

Micrometeorology, by O. G. Sutton. Pp. xii + 333. (London: McGraw-Hill, 1953.) 61*s*.

Micrometeorology deals with the physical processes in the lowest layers of the earth's atmosphere, mainly in the first hundred metres. It has, therefore, important applications in the fields of agriculture, biology, hydrology and atmospheric pollution. The appearance of Professor Sutton's book marks an important event in that it presents the first complete textbook treatment of the subject. Geiger's *Das Klima in der bodennahen Luftschicht* has, of course, been a standard work for 25 years, but it is mainly descriptive and does not attempt to develop the subject in a formal, quantitative manner. In view of his own important contributions it is fitting that Professor Sutton should fill this gap in the meteorological literature, which he has done with the elegance and style we associate with his writings.

The first five chapters of the book deal with the basic physics and mathematics required for the study of the processes occurring in the earth's boundary layer, under the following headings: The atmosphere at rest, Laminar flow in the atmosphere, Turbulent flow, Heat transfer and problems of diffusion, and Radiation. The last three chapters deal with the temperature field, the wind field and the processes of diffusion and evaporation in the lowest layers.

The presentation is everywhere concise, clear and polished, and the author has done his best to bring out the physical interpretation of the mathematical arguments—not an easy task in this subject. But it is in the very authoritativeness and neatness of the treatment that the reviewer senses a certain danger; the student or the newcomer to the subject may feel that it is much better understood and based on firmer theoretical foundations than is really the case. He may be lulled into a sense of false security at the outset by the author's remark in the preface, that micrometeorology presents a less forbidding aspect to the newcomer than the more familiar branches of meteorology. True, the prospective experimenter will have the advantage of working in the most accessible region of the atmosphere where he can more easily amass instrumental records, but, in the absence of anything like a satisfactory theory of atmospheric turbulence and radiative transfer, he will have a hard time in interpreting them without recourse to the usual semi-empirical approach. This state of affairs becomes obvious, of course, to the experienced meteorologist in reading Professor Sutton's book, but it might be kinder to warn the uninitiated at the outset. However, a book must be largely judged in the light of the author's aims and there can be little argument that Professor Sutton has produced a most valuable and readable work of reference on a difficult subject. The research worker will be grateful for this summary of the widely-dispersed literature and will find the bibliographies given at the end of each chapter most useful.

There appear to be very few errors, misprints etc., but *macro*meteorologists will not take kindly to the author's assertion at the bottom of p. 18 that, in the free atmosphere, the motion of the air may, to a high degree of approximation, be considered free from vorticity, nor to his subsequent remarks about the generation of vorticity.

The book is excellently produced but the price of 61*s*. (only 44*s*. in the U.S.A. !) may act as a deterrent to many who otherwise would and should have it on their bookshelves. B. J. MASON.

CONTENTS OF SECTION A

PAGE

Dr. R. E. B. Makinson and Mr. J. S. Turner. On Perturbation and Variation Methods 857

Mr. J. S. Turner and Dr. R. E. B. Makinson. A Comparison of Various Methods of Solving the Central Force Scattering Problem 866

Prof. H. S. Green. A Pre-Renormalized Quantum Electrodynamics 873

Mr. K. Parker. The Effect of Nuclear Multipole Moments on Electron Scattering 881

Dr. R. F. Barrow, Dr. G. Drummond and Mr. H. C. Rowlinson. The Absorption Spectrum of SnS Vapour in the Ultra-Violet and Schumann Regions . . 885

Dr. E. P. Wohlfarth. The Energy Band Structure of a Linear Metal . . 889

Dr. T. C. Griffith. The Scattering of 10 and 14 mev Neutrons by Deuterons . 894

Dr. B. H. Bransden and Dr. A. Dalgarno. The Calculation of Auto-Ionization Probabilities—I : Perturbation Methods with application to Auto-Ionization in Helium 904

Dr. B. H. Bransden and Dr. A. Dalgarno. The Calculation of Auto-Ionization Probabilities—II : A Variational Method for Radiationless Transitions with Application to the (2s)^2 ^{1}S–(1sks)^{1}S Transition of Helium 911

Prof. J. B. Birks and Mr. A. W. Little. Photo-Fluorescence Decay Times of Organic Phosphors 921

Prof. W. Heitler and Dr. Ch. Terreaux. Interpretation of Cosmic Ray Jets . 929

Research Notes :

Dr. D. K. Butt. The Efficiency of the Anthracene Scintillation Counter . 940

Mr. H. S. Murdoch. The Half-Life of ^{181}Ta^{m} and the Delayed Coincidence Method 944

Dr. S. Raimes. The Compressibility of Metallic Aluminium 949

Letters to the Editor :

Mr. H. Albers-Schönberg, Mr. K. Alder, Mr. E. Heer, Dr. T. B. Novey and Dr. P. Scherrer. The Influence of Combined Electric and Magnetic Interaction on Gamma–Gamma Directional Correlation 952

Dr. D. Bijl and Mr. A. C. Rose-Innes. Temperature Change in the Paramagnetic Resonance Spectrum of Copper Lanthanum Nitrate 954

Dr. W. G. Davey and Prof. P. B. Moon. The Resonant Scattering of ^{198}Hg Gamma Rays 956

Dr. T. A. S. Jackson. Accidental Degeneracy of Hydrogen 958

Reviews of Books 959

Contents of Section B 960

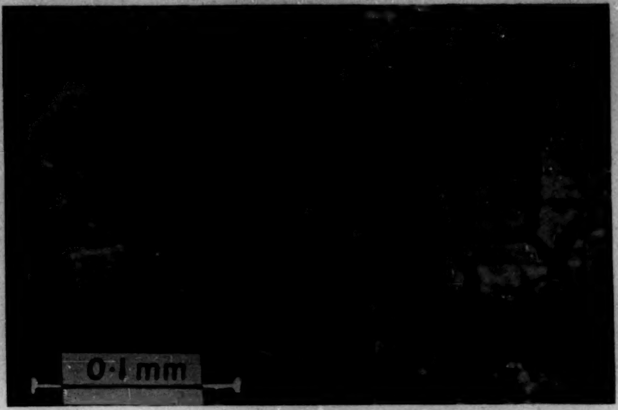


PLATE I.

Flow of 0·05%C steel at 1106°C.

Thermally etched surface showing

(*a*) discontinuities in scratch marks at austenite grain boundary, indicative of grain boundary slip, and

(*b*) coarseness of grain boundaries running perpendicularly to the specimen axis.

($\sigma = 158$ kg cm^{-2}; $d\epsilon/dt = 7\cdot5 \times 10^{-5}$ sec^{-1})

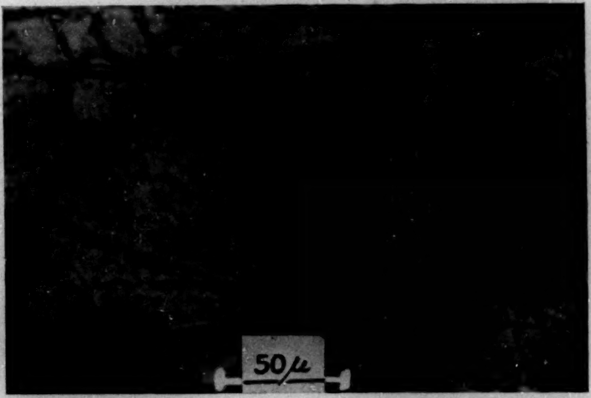


PLATE II.

Flow of 0·23%C steel at 1000°C.

Thermally etched surface showing

(*a*) curvature in originally straight longitudinal scratch marks, indicative of intragranular deformation, and

(*b*) coarseness of boundaries running perpendicularly to specimen axis.

The dark spot on the right of the central grain is part of a ferrite grain.

($\sigma = 294$ kg cm^{-2}; $d\epsilon/dt = 1\cdot8 \times 10^{-4}$ sec^{-1})

The Origin of Stress in Metal Layers Condensed from the Vapour in High Vacuum

By H. P. MURBACH and H. WILMAN

Applied Physical Chemistry Laboratory, Imperial College, London

MS. received 9th July 1953

Abstract. Measurements are made of the stress in deposits of Ni, Fe, Pd, Au, Cu, Ag, Sb, Al, Bi, Mg and Zn condensed from the vapour in high vacuum on copper strips used as substrates. The stress is found to be characteristically of a tensile nature and the amount of stress is different for different metals.

The occurrence, nature and order of magnitude of the stress are explained as arising from the contraction of the upper regions of the growing deposit as they cool after their deposition, which takes place at a temperature generally several hundred degrees higher than the initial substrate temperature. According to this interpretation, this contraction will begin to set up a tensile stress in any stratum of the deposit as soon as the temperature of the region concerned falls below the recrystallization temperature of the deposit metal. An estimation of the stress to be expected from this cause falls in the region of the experimentally determined values.

In Al a compressive stress is observed if the metal is evaporated in an imperfect vacuum, the air pressure being about 10^{-4} mm Hg or higher. Evidently appreciable oxidation of the aluminium occurs and the compressive stress is associated with the formation of the oxide between the metal crystals or included to some extent within the crystals. Results for electrodeposits are analogous.

§ 1. Introduction

The origin of stresses in electrodeposits has been much investigated, but with little success in accounting for the nature of the stress (tensile or compressive) or the amount of stress. Recent work in this laboratory (Finch, Wilman and Yang 1947, Finch and Layton 1951), has shown how the main factors influence the crystalline structure of electrodeposits, and it was thought that it might also be possible to correlate these factors and the deposit structure with the stress in the deposits. Experiments made to test this possibility indicated the desirability of first investigating whether, in the simpler case of deposits condensed from the vapour *in vacuo*, similar stresses occur and can be explained. Although some experiments of this kind have been made recently (Crittenden and Hoffman 1950, Hoffman, Anders and Crittenden 1953), these were only with one metal (nickel), and thus allowed no general conclusions to be drawn as to the origin of the stress in condensed metal deposits.

Measurements made on deposits of many metals condensed *in vacuo* are now described below, and show that characteristically tensile stresses are indeed observed, of magnitudes differing widely from one metal to another. The values observed lead to the conclusion that the stress is due to contraction of the upper regions of the growing deposit as they cool after their formation, which takes place at a temperature generally several hundred centigrade degrees above the substrate temperature.

§ 2. EXPERIMENTAL PROCEDURE

Metal deposits were condensed *in vacuo* on one side of a copper strip which was clamped at one end, so that the magnitude of the stress in the deposit could be calculated from the observed deflection of the free end of the strip. This method has been much used in the study of stresses in electrodeposits and was first used by Stoney (1909). Brenner and Senderoff (1949) have recently discussed the derivation of the formula used to calculate the stress S (kg cm^{-2}) from the deflection δ (cm) and confirmed Stoney's equation

$$S = Et^2\delta/3dl^2 \qquad \ldots\ldots(1)$$

where t is the thickness in cm of the substrate strip, E is its Young's modulus, and l is the length in cm of the strip carrying the deposit of thickness d cm. This assumes that the deposit thickness is negligible compared with the substrate thickness.

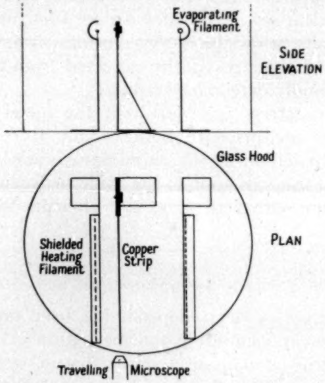


Fig. 1. Diagram of apparatus used for determining stress in condensed metal layers.

The copper strips used as substrates were 3·4 cm long, 0·97 cm wide and 0·0155 cm thick. To smooth the surface and avoid possible influence of the crystal orientation by the substrate, the strips were polished on a polishing wheel and then, immediately before use, with 'Bluebell' metal polish applied on cotton wool. The strips were washed in water, acetone, and finally benzene, to ensure freedom from surface contamination, especially grease. The Young's modulus E of such strips was determined by observing the deflection of the free end when the strip was clamped at one end with the strip face horizontal, and a 0·5 g weight was hung on the free end. In this way E was found to be $6{\cdot}5 \times 10^5$ kg cm^{-2}.

The copper strip was next clamped at one end with its length horizontal, and its face vertical, in a high-vacuum unit, and an electrically heated tungsten coil 5 cm away was used to evaporate any desired metal to form a deposit on the strip. The position of the free end of the strip was observed through the glass wall of the vacuum chamber, by means of a travelling microscope, to an accuracy of 0·001 cm. Figure 1 shows the arrangement diagrammatically, and also the disposition of an auxiliary tungsten coil behind the copper strip, when tests were made to find what stresses were developed in deposits formed on substrates at elevated temperature. Estimates of the mean substrate temperature during and after a deposition were made by use of a thermocouple in close contact with the back face of the strip.

Deposits about 1000–3000Å thick were prepared of Ni, Fe, Pd, Au, Cu, Ag, Sb, Al, Bi, Mg and Zn, the time of deposition being of the order of 0·5 minute. The deflections observed were mostly between 0·01 and 0·05 cm. Blank experiments, in which the front face of the substrate was heated by radiation from the evaporator spiral alone without evaporating any metal from it, showed negligible deflection of the copper strip.

The structure of the deposits was investigated by electron diffraction in a Finch-type camera (Finch and Wilman 1937), using 50–60 kv electrons and a camera length of 49 cm.

§ 3. RESULTS

The stresses observed for the metals investigated are listed in table 1, where the metals are arranged in order of decreasing stress. For comparison the melting points and the temperatures of recrystallization or 'recovery' (of mechanical properties after cold working) are also given in table 1, and it will be seen that these properties also are in roughly the same sequence of decreasing values.

Table 1. Observed Stresses in the Metal Deposits Condensed from the Vapour *in vacuo*, and some Associated Properties

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fe	2	3100	1539	350–450[1,2]	12	21	200–300	~3000*
Ni	as 4	3500	1455	530–660[1,2]	14	21	380–500	~3000*
Pd	3	1400	1555	450[3]	11	11	300	~1500*
Cu	4	900	1083	200–230[1,2]	17	12	80–110	~1000*
Au	as 4	850	1063	200[1,2]	14	8	80	900
Ag	as 4	750	960	200[1,2]	19	8	80	1200
Sb	—	250	630	—	11	8	—	—
Al	5, 6	100	659	150[1]	24	7	30	500
Mg	7	0	651	150[2], 200[1]	27	4	30	300
Bi	8	0	271	—	13	9	0?	0?
Zn	9	0	419	< 20[1]	26	9	0	0

(1) metal; (2) figure no.; (3) approximate mean observed stress (kg cm^{-2}); (4) melting point (°c); (5) recrystallization temperature (°c); (6) $10^6\alpha$ [coefficient of linear expansion]; (7) $10^{-11}E$ (dyn cm^{-2}); (8) approximate temperature range ΔT (°c) in which stress is developed; (9) estimated theoretical stress (kg cm^{-2}).

* Above the elastic limit.

[1] Chalmers 1951. [2] Masing 1950. [3] Tamman and Dreyer 1933.

This suggests that the stress is in some way related to the degree of mobility of the atoms in the deposit at the substrate temperature or the temperature of the upper regions of the growing deposit.

The electron diffraction patterns from the deposits (figs. 2–9 (Plate)) showed that for Ni, Fe, Pd, Cu, Au, Ag and Al the crystals were small (~100–150 Å diameter) and randomly disposed or weakly orientated with a common axis inclined away from the substrate normal towards, but not as far as, the direction of the vapour stream. On the other hand for Sb, Mg, Bi and Zn the crystals were much larger (>400 Å diameter) and were orientated, in the mean, with a net plane parallel to the substrate or practically so. In this last case the one-degree orientation tended to be limited further in azimuth, as a result of pronounced plate-like crystal habit, which caused preferential growth of those crystals having their main surface more or less facing the source of metal vapour.

It has recently been shown (Evans and Wilman 1952) that these two types of orientation are associated respectively with low and high atomic mobility of the deposited atoms on the substrate and the growing deposit. The melting point of the metal was shown to be a rough inverse measure of the degree of mobility of the atoms on the substrate and the deposit surface at the temperature reached during the deposition, with the substrate initially at room temperature. The above electron diffraction data in the present experiments thus also confirm that the stress developed in a metal deposit tends to be larger when atoms in the deposit have low mobility during the formation of the deposit.

From these general conclusions it would be expected that deposits formed on substrates which were initially at raised temperature would have less stress than similar deposits formed on substrates initially at room temperature. To test this, several pairs of deposits of nickel were prepared simultaneously on two copper strips, one of which was initially at room temperature and the other at about 125°c (heated by radiation from a tungsten coil behind the strip, as in fig. 1). Table 2 shows typical results from two experiments where the same two copper substrate strips were used, but interchanged in the two experiments. It will be seen that the stress in the deposit on the heated substrate was reduced in both cases by about 25%, thus confirming further the relationship between the stress and the temperature of formation of the deposit, and hence the atomic mobility in the deposit.

Table 2. Stresses Observed in Nickel Condensed from the Vapour *in vacuo* on Polished Copper Substrates at Room Temperature and at 125°c respectively (assuming same E as at Room Temperature)

Deposit thickness (Å)	1250	630
Stress (strip initially at room temp.): S_1 (kg cm^{-2})	3720	4590
Stress (strip initially at 125°c): S_2 (kg cm^{-2})	2840	3410
S_2/S_1 (as %)	76·5	74·4

Although the above results for the metals investigated are consistently in agreement with this conclusion, wide variations in the stress occurred in the case of aluminium when deposits were prepared in residual air in the vacuum chamber at a pressure higher than about 10^{-4} mm Hg. In this case compressive stresses up to several hundred kg cm^{-2} were observed, the stress being higher the higher the residual air pressure above this value, up to about 10^{-3} mm Hg, when it was difficult to evaporate the aluminium owing to oxidation of its surface. The electron diffraction patterns from the aluminium deposits correspondingly became less and less clearly arced until at the highest observed compressive stresses (highest residual air pressure) the patterns consisted of rings due to randomly orientated crystals. These profound changes in the deposit structure and stress must be attributed to appreciable oxidation of the aluminium during deposition. No noticeable rings or haloes due to aluminium oxide occurred in the diffraction pattern, however, though it would be difficult to detect clearly the presence of even a considerable proportion of amorphous oxide in presence of the crystalline aluminium.

§ 4. DISCUSSION AND CONCLUSIONS

The above results on deposits of eleven metals show clearly that there is a systematic relationship between the stress in the deposits and the degree of mobility of the atoms in the upper regions of the growing deposit. We therefore conclude that the observed tensile stress is due to the cooling of the deposit after its formation at a surface temperature which must be several hundred centigrade degrees (cf. Evans and Wilman 1952), owing to receipt of radiation from the hot evaporator filament, and energy carried by the deposited atoms. Thus, when the metal crystals are growing at the deposit surface by addition of the arriving atoms they are thermally expanded. As these layers become covered by further growth of the deposit, their temperature must decrease progressively in accordance with the temperature gradient in the deposit, and the crystals must contract. As they do so, appreciable atomic migration persists and must cause readjustments at and near the grain boundaries, i.e. recrystallization, until the temperature falls below the recrystallization temperature, and thereafter the further cooling must cause a tensile stress to be developed throughout the deposit, as is indeed observed. The amount of stress developed must therefore depend on the difference between the final deposit temperature and the recrystallization temperature, or the highest temperature reached if this is less than the recrystallization temperature.

Table 1 shows the linear coefficient of thermal expansion α (for approximately the range of temperature concerned) and the Young's modulus E of the metals investigated. It is seen that the metals which developed the highest stress have a high recrystallization temperature and also often a high Young's modulus, in agreement with expectations.

The approximate mean stress to be expected according to the above conclusions can be estimated as follows. For a temperature drop ΔT from the recrystallization temperature to the final substrate temperature, the relative contraction of the deposit would be $\Delta l/l = \alpha\Delta T$. The stress caused, if the layer is not allowed to contract, owing to the forces exerted on it by the substrate, is thus

$$S = E(\Delta l/l) = E\alpha\Delta T. \qquad \ldots\ldots(2)$$

From the values for E, and the roughly estimated temperature drop ΔT, the theoretical order of stress expected was estimated and is given in table 1. For Fe, Ni, Pd and Cu the contractions impose stresses which fall outside the range in which the strain is proportional to the stress, hence the stress was roughly estimated from the available data.

It is seen that the stresses estimated theoretically in the above way agree well in order of magnitude with the observed values. Thus, although much previous work on stress in electrodeposits has hitherto not led to any clear understanding of the origin and amount of the stress, the present experiments, and our interpretation of them, show that at least the stresses developed in metal deposits condensed from the vapour can be explained satisfactorily both with respect to their tensile nature and the amount of the stress and also the correlation with the atomic mobility occurring during the crystal growth. It seems likely that the stress in electrodeposits may be explained by a somewhat similar process, involving (in the absence of disturbing effects) formation of the crystals in the deposit surface in an initially expanded state, followed by their contraction.

Furthermore, in the case of aluminium the observations in relation to the pressure of the residual air in which the evaporation was carried out indicate that the compressive stresses are associated with appreciable oxidation of the metal during growth of the deposit. This strongly suggests that in electrodeposits the compressive stresses sometimes observed may arise by chemical reactions or co-deposition processes leading to occlusion or inter-crystalline adsorption of foreign materials in the metal deposit. Experiments made by the authors have in fact already shown, in copper and nickel electrodeposits on copper substrate strips, that when a compressive stress was developed the electron diffraction patterns usually showed, in addition to metal diffraction rings or arcs, other diffractions indicating presence of co-deposited material. It is hoped to carry out further experiments on stress in electrodeposits with a view to extending in more detail these general conclusions.

A further point clarified by the present results is the question of the relation between stress and orientation. Bozorth (1925) and Wyllie (1948) considered that the preferred orientation often found in electrodeposits can be explained as arising as a ‘ drawing texture ’, from plastic deformation occurring in the deposit when the stress exceeds the elastic limit of the metal. The factors promoting orientated crystal growth in electrodeposits have been elucidated by Finch, Wilman and Yang (1947), and Yang (1948) in this laboratory has also shown by combined electron diffraction and stress investigations, that this theory does not account for the orientations observed. The present results show the analogous conclusion for deposits condensed from the vapour *in vacuo*, i.e., that the observed orientations, where present, are not those expected from Bozorth's theory and are often very weak or absent (for example, some iron deposits investigated were randomly orientated) even when stresses are present much exceeding the elastic limit. It seems that when the stresses in the deposits exceed the elastic limit the plastic deformation which occurs is more likely to have a disorientating effect, in the case of deposits which initially grow in a preferred one-degree orientation relative to the substrate.

REFERENCES

BOZORTH, R. M., 1925, *Phys. Rev.*, **26,** 390.

BRENNER, A., and SENDEROFF, F., 1949, *Bur. Stand. J. Res., Wash.*, **42,** 105.

CHALMERS, B., 1951, *The Structure and Mechanical Properties of Metals* (London : Chapman and Hall), p. 70.

CRITTENDEN, E. C., and HOFFMAN, R. W., 1950, *Phys. Rev.*, **78,** 349.

EVANS, D. M., and WILMAN, H., 1952, *Acta Crystallogr.*, **5,** 731.

FINCH, G. I., and LAYTON, D. N., 1951, *J. Electrodepositors' Tech. Soc.*, **27,** paper no. 9.

FINCH, G. I., and WILMAN, H., 1937, *Ergebn. exakt. Naturw.*, **16,** 353.

FINCH, G. I., WILMAN, H., and YANG, L., 1947, *Discussions of the Faraday Society*, **1,** 144.

HOFFMAN, R. W., ANDERS, F. J., and CRITTENDEN, E. C., 1953, *J. Appl. Phys.*, **24,** 231.

MASING, G., 1950, *Lehrbuch der allgemeinen Metallkunde* (Berlin : Springer), p. 431.

STONEY, G., 1909, *Proc. Roy. Soc.* A, **82,** 172.

TAMMANN, G., and DREYER, K. L., 1933, *Ann. Phys., Lpz.*, **16,** 111.

WYLLIE, M. R. J., 1948, *J. Chem. Phys.*, **16,** 52.

YANG, L., 1948, *Thesis*, University of London.

The Variable-Radius Semicircular Magnetic Focusing β-Ray Spectrometer

By C. G. CAMPBELL and J. KYLES
Department of Natural Philosophy, University of Edinburgh

Communicated by N. Feather; *MS. received* 26*th May* 1953

Abstract. Li's method for determining the line shape in a variable-radius spectrometer is discussed, and a rapid approximate method is given. Methods of normalization for the continuous β-particle spectrum and for conversion electron lines are presented. The advantages and disadvantages of this type of spectrometer are discussed.

§ 1. Introduction

THE semicircular focusing β-ray spectrometer was designed primarily for the photographic determination of the $H\rho$-values of internal conversion electron lines. More recently β-particle counters have been used as detectors instead of the photographic plate, and the double spectrometer, first proposed and constructed by Feather (1940) and now equipped with scintillation counters and short resolving time coincidence circuits, has proved a useful instrument for the determination of certain features of the energy level scheme of the daughter body in β-decay (Kyles, Campbell and Henderson 1953). Each half of this instrument constitutes a fixed-field variable-radius spectrometer, and it is felt that an examination of its performance would be useful in the light of our experience with it.

An inherent disadvantage of this type of spectrometer lies in the small angle for collection of the electrons in the plane of focusing, when a resolution* approaching 1% is required. On the other hand, changes in the width of the slit selecting the sheaf of β-particles to be focused may be very simply made, and the width of the entrance slit of a scintillation counter, which requires no window, may be varied at will. The instrument is therefore extremely versatile, and permits the adoption of geometrical conditions to suit the particular experiment to be performed.

To ascertain these conditions theoretically a study of the structure of a β-ray line shape is necessary.

§ 2. Li's Method

The basis of the method adopted was given by Li (1937). In this method three sets of coordinates, illustrated in fig. 1, are used. The coordinates ξ, η, ζ describe the source-space, x, y, z the source-slit space, and x', y', z' the image space for the particular line under consideration. Let $\mathrm{AB}=2s$, $\mathrm{OS}=d$, and for the line in question let $\mathrm{PO}=a$, $\mathrm{PS}=2\rho$.

It is then shown by Li that the intensity at the point $(x', 0)$ of the image plane is given by dividing the source plane into three domains and integrating, in turn, over those portions of these domains occupied by the source, a function K

---

* We define resolution as $\Delta p/p$ where Δp is the width at half height of a monoenergetic line of momentum p. Resolving power is the reciprocal of this quantity.

which is weighted with the values 0, 1 or 2 according to the particular domain chosen. This function K is given by

$$K=\frac{n}{4\pi^2\rho(2a)^{1/2}}\frac{1}{(x'-\xi-\mu)^{1/2}} \qquad \ldots\ldots(1)$$

in which n represents the number of electrons leaving unit area of the source per second and $\mu=2\eta^2/\pi^2a$.

The boundaries of the domains in the source plane are shown to be given by

(i) The parabola $$x'-\xi=2\eta^2/\pi^2a. \qquad \ldots\ldots(2)$$

This is the locus of points in the source from which electrons reach the image point, having passed through the line $x=x'$ of the source slit. $(x', 0)$ is the apex of this parabola in the source coordinates.

(ii) The ellipse 'B' $$x'-\xi=\frac{2\eta^2}{\pi a^2}+\frac{2\rho^2}{ad^2}(s+\xi)^2 \qquad \ldots\ldots(3)$$

which is determined by those trajectories which graze the edge B of the source slit nearer to the image point $(x', 0)$.

(iii) The ellipse 'A' $$x'-\xi=\frac{2\eta^2}{\pi^2 a}+\frac{2\rho^2}{ad^2}(s-\xi)^2 \qquad \ldots\ldots(4)$$

which is determined by those electron trajectories which graze the edge A of the source slit remote from the image point.

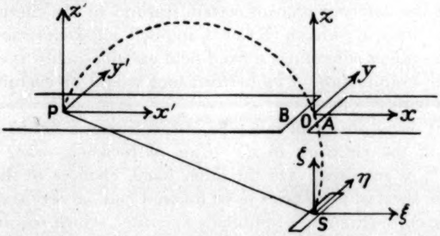


Fig. 1.

The coordinates of the centre of symmetry of the ellipse 'A', which remains collapsed in its centre until $x'>s-ad^2/8\rho^2$, are $(-(ad^2-4s\rho^2)/4\rho^2,\ 0,\ 0)$ in the source frame. For small radii of curvature this centre lies to the left of the source (in the sense of fig. 1), and, so long as $a>d$, which is the case in most spectrometers, as ρ increases moves steadily to the right, passing through the origin of the source plane coordinates when $\rho^2=ad^2/4s$. The centre of symmetry of the ellipse 'B', which lies $2s$ to the left of that of 'A', remains to the left of S throughout.

Three distinct cases have to be examined, corresponding to $\rho^2<, =, >ad^2/4s$, respectively.

Case 1. $\rho^2<ad^2/4s$.

The diagrams of fig. 2 show the general character of the source domains as the image point x' traverses the X′ axis of the image space. The apex of the parabola moves with the same speed as x'. The axes of the ellipses increase with x', but their centres of symmetry remain at rest and their eccentricities constant.

Source domains of zero, single and double weight are indicated on the diagrams.

Case 2. $\rho^2 = ad^2/4s$.

Figure 3 illustrates the types of source domain encountered here. The ellipse 'A' does not open until $x' = s/2$ in this case.

Case 3. $\rho^2 > ad^2/4s$.

Figure 4 illustrates the types of source domain encountered with these radii of curvature.

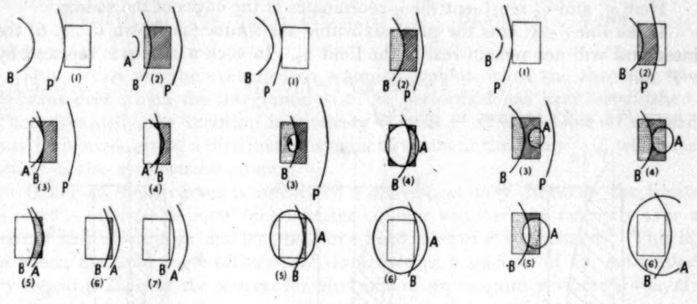


Fig. 2. $\rho^2 < ad^2/4s$. Fig. 3. $\rho^2 = ad^2/4s$. Fig. 4. $\rho^2 > ad^2/4s$.

Typical source domains for various image points (see text). Not drawn to scale.
A, ellipse 'A'; B, ellipse 'B'; P, parabola.
▨ source domain of weight 2; ░ weight 1; ☐ weight 0.

In all cases source domains lying between the following limits contribute (with single weight) to the intensity of the image point. The overlapping of two domains (it is clear, physically, that not more than two domains can overlap) automatically gives those regions double weight.

(i) From the left-hand side of ellipse 'A' to the left-hand side of ellipse 'B'.
(ii) From the right-hand side of ellipse 'B' to the parabola.*
(iii) From the right-hand side of ellipse 'A' to the parabola.*

So far our constructions are relevant only to calculations of intensity at image points along the X′ axis. To extend the discussion to include points along the image lines $y' = \pm c$, where c is a constant, it is convenient to regard the source as being displaced a distance $\mp c$ along the η-axis, and then to consider boundaries of the source domains appropriate to points along the X′ axis of the image.

§ 3. ANALYTIC EXPRESSION FOR THE INTENSITY AT AN IMAGE POINT

The intensity at an image point may now be written as

$$\Phi = 2\int\int K\,d\xi\,d\eta + \int\int K\,d\xi\,d\eta$$

where the first term is integrated over the source domain of double weight and the second term over the domain of single weight.

---

* If $x' > s$ then the right-hand side of ellipse 'A' replaces the parabola in (ii), and (iii) does not contribute.

Only when n is constant (as is assumed throughout our discussion) and a domain extends from a straight line (say $\xi=\xi'$) to a parabola may the integration be performed analytically. Such a hypothetical domain is illustrated in fig. 5.

Integration in this case gives

$$I=F(v,\,\eta_2)-F(v,\,\eta_1) \qquad \ldots\ldots(5)$$

in which

$$F(v,\,\eta)=\frac{Gv^2}{P}\left\{\operatorname{arc\,sin}\frac{P\eta}{v}+\frac{P\eta}{v}\left(1-\frac{P^2\eta^2}{v^2}\right)^{1/2}\right\} \qquad \ldots\ldots(6)$$

$v^2=x'-\xi'$, $P^2=2/\pi^2a$ and $G=n/4\pi^2\rho(2a)^{1/2}$.

Here η_1 and η_2 represent the η-coordinates of the edges of the source.

If the line $\xi=\xi'$ cuts the parabola within the source (as shown in fig. 6) the integrand will not remain real at the limit η_2. In such a case η_2 is replaced by v/P and

$$I=F(v,\,v/P)-F(v,\,\eta_1) \qquad \ldots\ldots(7)$$

where $F(v,\,v/P)$ reduces to $\frac{1}{2}\pi Gv^2/P$.

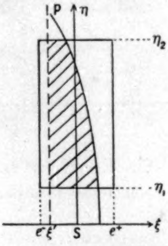


Fig. 5 (see text).

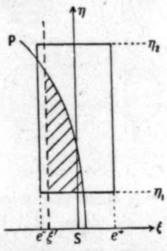


Fig. 6 (see text).

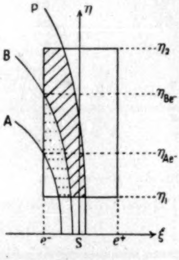


Fig. 7 (see text).

The domain of integration does not usually extend from a straight line to a parabola. In all cases, however, the integral over the correct domain may be obtained by appropriate sums or differences of integrals extending from suitably chosen straight lines, or from one or both ellipses, to the parabola, an appropriate choice for the limits for η being made for the various integrals. An example is given below.

It remains therefore to show how K may be integrated over a region of the source extending from an ellipse to a parabola. The integration with respect to ξ may be performed analytically, but an approximate method is required for the integration with respect to η. Simpson's rule is suitable for this numerical integration, since it is accurate for functions with constant third differences, and this integrand possesses almost constant second differences. An evaluation with three ordinates only gives accuracy to four significant figures.

For integration from ellipse 'A' to the parabola the integral takes the form

$$I=2G\int_{\eta_1}^{\eta_2}\left\{x'+\frac{Q^2}{2P^2}-s-P^2\eta^2\mp(\alpha_{\mathrm{A}}{}^2-Q^2\eta^2)^{1/2}\right\}^{1/2}d\eta \qquad \ldots\ldots(8)$$

in which $Q=d/\pi\rho$ and $\alpha_{\mathrm{A}}=(d/4\rho^2)[a\{ad^2-8\rho^2(s-x')\}]^{1/2}$, which is the minor axis of the ellipse in question.

The negative and positive signs give integrations from the right-hand edge and the left-hand edge of the ellipse respectively.

The corresponding integrals will, for brevity, be written as

$$I=\int_{\eta_1}^{\eta_s} A_{\mathrm{R}}\,d\eta \quad\text{and}\quad I=\int_{\eta_1}^{\eta_s} A_{\mathrm{L}}\,d\eta. \qquad \ldots\ldots(9)$$

For integration from ellipse 'B' to the parabola s is replaced by $-s$, and B is used in place of A.

To illustrate the manner in which the intensity at a selected image point may be obtained in a practical case, let the source extend from e^- to e^+ in the ξ-direction; then for the domains illustrated by fig. 7, Φ is given by

$$\Phi=2F(v,\,\eta_2)-F(v,\,\eta_{\mathrm{B}e^-})-F(v,\,\eta_{\mathrm{A}e^-})+\int_{\eta_1}^{\eta_{\mathrm{B}e^-}} B_{\mathrm{R}}\,d\eta+\int_{\eta_1}^{\eta_{\mathrm{A}e^-}} A_{\mathrm{R}}\,d\eta$$

where $v^2=x'-e^-$.

Other cases may be evaluated in a similar fashion when the shape of the domains over which the integration is to be performed has been established. Thus, eventually, the variation in intensity Φ with y' at each value of x' used may be plotted, giving a distribution symmetrical about the line $y'=0$, when the source is also symmetrical about $\eta=0$.

If one of these curves is integrated with respect to y' between the limits $+l$ and $-l$ (where $2l$ is the length of the counter window) the intensity over a narrow strip of width dx' and length $2l$ for a fixed value of x' is obtained. This is, in effect, the solid angle $(d\Omega/dx')\,dx'$, expressed as a fraction of 4π, subtended by the strip $2l\,dx'$ at the source, for electrons of momentum p where $\rho=pc/He$ (ρ being the radius of the trajectory in a plane at right angles to the magnetic field H).

The integration may be carried out by a suitable adaptation of Hardy's '37' numerical integration formula.

The line profile $S(x')$ may now be found by numerical integration. If $2w$ is the width of the counter window we have

$$S(x')=\int_{x'-w}^{x'+w}\left(\frac{d\Omega}{dx'}\right)dx'.$$

§4. AN APPROXIMATE METHOD FOR FINDING THE LINE PROFILE

The arithmetical work involved in the calculation of the line profile by the method outlined is lengthy, and for many purposes a less exact but more rapid method is desirable. The variation in intensity along lines parallel to the X′ axis of the image coordinates follows the same general shape for all values of y' used in practical cases. Thus the change in shape of a single member of this set of curves is sufficient to indicate the effect on solid angle and on resolution produced by a change in the geometry of the spectrometer. Attention will therefore be paid to the approximate determination of the intensity distribution along the X′ axis only.

The following simplifications will then be made: (i) it will be assumed that the function K remains sensibly constant for all points of the source, (ii) the curvature of the ellipses and of the parabola over the area of the source will be neglected.

The first assumption implies that the intensity at an image point $(x', 0)$ is now determined simply by the areas of the domains in the source, suitably weighted. The second assumption implies that, since the source is assumed to be rectangular in shape, the areas of the domains are proportional to the distances between the straight lines in the source plane which now replace the parabola and the ellipses. Clearly this second assumption will not apply to those cases in which $\rho^2 \sim ad^2/4s$, for then the centre of ellipse 'A' will lie in, or near, the source. The centre of symmetry of this ellipse is at $\xi = s(r-1)/r$ and the ellipse is real for $x' > s(2r-1)/2r$, where $r = 4s\rho^2/ad^2$.

A diagram is constructed with axes ξ and x', as illustrated in fig. 8 (*a*). The lines $\xi = \pm e$ indicate the extent of the source in the ξ-direction. The movement of the apex of the parabola through the source, corresponding to motion of the point x' along the X′ axis of the image, is obtained by setting $\eta = 0$ in eqn. (2), giving the line $x' = \xi$. This line is marked PP′.

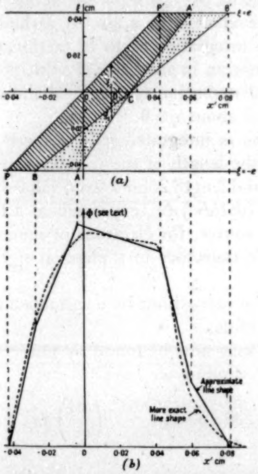


Fig. 8. (*a*) see text. (*b*) comparison of line shape given by the rapid approximate method with that obtained by the more exact method of §§ 2 and 3.

The movement of the extremity of the minor axis of the ellipse 'A' is given by putting $\eta = 0$ into eqn. (4), which becomes

$$x' - \xi = \frac{2\rho^2}{ad^2}(s-\xi)^2. \qquad \ldots\ldots(10)$$

On the (x', ξ) diagram this is a parabola.

The ellipse 'B' gives a second parabola (with s replaced by $-s$), intersecting the first on the x' axis at $x' = 2\rho^2 s^2/ad^2$ (denoted by C).

These parabolae are marked ACA′ and BCB′ respectively. Their curvature is small, and they may usually be replaced by pairs of straight lines, AC and CA′ and BC and CB′, the x'-coordinates of A and A′ being obtained by solving eqn. (10), with $\xi = \pm e$; similarly for B and B′.

From this diagram the desired variation with x' in the width of the source domains may be obtained immediately, since the effective areas of the source are proportional to the separations of the lines in the ξ-direction. The shading indicates regions of double and single weight. Hence the intensity corresponding to the point D, say, is proportional to $\xi_1 + 2\xi_2$ as shown in the figure.

The approximate variation in intensity with x' obtained in this way is shown in fig. 8 (*b*), along with the curve obtained by the method of §§ 2 and 3 for comparison. The more exact curve has been normalized to the height of the approximate curve.

The approximate method will be seen to give directly a close approximation to the spread of intensity along the X′ axis, but the intensity itself is not given absolutely.

For sources which lie in the (ξ, η) plane, relative intensities, at various radii, may be obtained by the approximate method by multiplying the ordinates by $1/\rho$, since the solid angle is, to a first approximation, proportional to $1/\rho$.

Sources set at an emission angle θ to the (ξ, η) plane may be similarly treated by projecting the source on to this plane at the angle made with the plane by that electron trajectory which leaves the centre of the source and, passing through the origin of the source slit plane coordinates, arrives at the origin of the image plane coordinates. Then when $2e$ is the projected width of such a source, the normalizing factor is replaced by $1/e\rho$.

The absolute intensity may always be obtained by calculating the value of the intensity at a single image point by the method of §§ 2 and 3.

§ 5. GENERAL REMARKS ON THE SPECTROMETER DESIGN

It is well known that the solid angle for collection of electrons in this type of spectrometer diminishes as the radius of curvature employed increases. On the other hand, the resolving power, defined for a constant field spectrometer as $\rho/\Delta\rho$, increases with ρ, but the rate of increase with ρ is less than linear.

In the example given it is clear that the whole source is not occupied by a domain of double weight for any image point. This condition is characterized by the flat top to the intensity distribution along the X′ axis of the image, caused by the movement, through a relatively broad source, of a narrow domain of double weight. An immediate increase in solid angle may therefore be effected by increasing the width of the source slit to make the x' coordinate of B (fig. 8 (*a*)) coincide with that of P′. The whole source then will contribute with double weight to the intensity of the image point in question at the radius of curvature chosen. The appropriate width of the source slit, given by solving eqn. (3) for s, with $\eta = 0$, $x' = e$ and $\xi = -e$, is then

$$s = \frac{d}{\rho}(ae)^{1/2} + e. \qquad \ldots\ldots(11)$$

At radii larger than that for which this condition applies it is found that the intensity distribution along the X′ axis of the image will again possess a flat top. The domain of double weight is now broader than the source. Resolving power is reduced, with no compensating improvement in solid angle. It is therefore advantageous to apply condition (11) to the case in which ρ is large; then at smaller radii the solid angle is rather less than the optimum value, but a compensating increase in resolving power is effected.

The approximate method outlined may be used to illustrate the variation in resolving power and in the solid angle with the position of the source. When the source is displaced towards the image by half its width (so that one edge now lies along the η axis of the source-space) the optimum source-slit width becomes

$$s=\frac{d}{\rho}(ae)^{1/2}+2e \qquad \ldots\ldots(12)$$

whereas displacement of the source away from the image by an equal amount leads to

$$s=\frac{d}{\rho}(ae)^{1/2} \qquad \ldots\ldots(13)$$

for the optimum source-slit width.

A detailed examination of these cases shows that the former source position gives increased solid angle, but smaller resolving power, whereas the latter gives a higher degree of resolution with a smaller solid angle; moreover this source position leads to an undesirable 'tail' on the low-energy side of the line profile.

The choice of the width of the counter slit is a matter of compromise, since the spread of the line profile increases considerably with ρ, the increase being more rapid as the source is displaced towards the image. The criterion adopted by us is that the width of the slit should be equal to the width of the distribution in intensity along the X′ axis at half the maximum height. The tendency then will be to obtain a detector slit rather narrower than the optimum width, since the spread in the line profile is rather greater than that in the distribution along the X′ axis, but this additional spread is small when compared with the changes in the half-width of the X′ axis distribution with ρ. For all source positions the point on the line profile which is representative of the radius ρ is given when $x'=0$.

§ 6. THE β-RAY CONTINUOUS SPECTRUM

Experimental observations of a β-ray continuous spectrum yield data in which the counting rate is given as a function of a, the distance of the centre line of the counter slit from that of the source slit. Corresponding to each value of a, the value of ρ may be calculated, and from the field H the $H\rho$-value corresponding to each detector setting may be deduced.

Only β-particles of a limited range of $H\rho$-values are focused by one field setting, and a complete mapping of the continuous spectrum usually involves several plots taken at different values of H so chosen that the $H\rho$-values of adjacent plots overlap.

For a given setting of the counter a very close approximation to the transmission factor, defined as the variation with momentum p of the probability of an electron entering the counter, is given by the line shape now plotted against p instead of a. This was shown to be the case by Lawson and Tyler (1940) for a spectrometer in which the source and the counter slit were coplanar, and remains true for this type of spectrometer since a is very nearly proportional to ρ over the small momentum range involved.

The counting rate at a given counter setting is the product of the area enclosed by the transmission factor curve and the rate at which electrons leave the source.

To obtain a true representation of that portion of a momentum spectrum which can be examined by movement of the counter at a fixed setting of the magnetic field, it is clear that the area enclosed by the transmission factor curves at various radii must be made constant by multiplication by an approximate factor dependent on ρ. This factor, by which the counting rate must be multiplied for normalization, will be referred to as the ρ-factor. To extend the normalization to different settings of the field, the observed counting rate, adjusted for variation of the transmission factor with ρ, must be divided by the magnetic field strength H as shown by Gurney (1925).

It has been shown by Lawson and Tyler (1940) that the 'representative' momentum of the distribution which enters the counter window at any particular setting is given by the abscissa of the centroid of the momentum distribution. It will be noted that this representative point differs from that of a conversion line.

§7. DETERMINATION OF THE ρ-FACTOR

The arithmetical labour involved in the calculation of the ρ-factor by the more accurate method of §§2 and 3 is prohibitive. Moreover, appropriate allowance cannot be made theoretically for such factors as the finite thickness (and fixed bevel) of the counter slit, which introduces scattering and absorption effects which are increasingly important for narrow slits. An experimental determination is therefore desirable and the following methods are available.

(i) The method outlined by Feather, Kyles and Pringle (1948) in which the counting rates are observed at a specific $H\rho$-value, chosen on a portion of a momentum spectrum of minimum slope, focused at different radii in different fields.

(ii) By observing the area enclosed by a given conversion electron line, again focused at different radii, and determining the factor necessary to make the areas constant.

(iii) By comparison of the spectrum observed for a given field setting with the same spectrum observed by other methods. Either a spectrum which has been well established using other types of spectrometer is required or a spectrum taken by using this spectrometer as a fixed radius instrument.

§8. THE CORRECT REPRESENTATION OF CONVERSION LINES

It can be shown that the method for normalizing the continuous spectrum is equally applicable to conversion lines provided that the conversion lines are plotted on the same momentum scale as that adopted for the continuous spectrum. The ratio of the area of a conversion line to that under the continuous spectrum gives the absolute intensity of the line. A detailed discussion of this, and allied topics, has been given by one of us (Campbell 1952).

§9. THE ADVANTAGES AND DISADVANTAGES OF THE VARIABLE-RADIUS SPECTROMETER

The following appear to be the only disadvantages of this type of spectrometer:

1. The solid angle, for a given resolution, compares unfavourably with that of a magnetic lens spectrometer.
2. Difficulty is experienced in screening the detector from unwanted γ-radiations, but in some cases this can be offset by the use of the proportional properties of scintillation counters.
3. An elaborate baffle-system to avoid counting scattered electrons cannot be incorporated.
4. The determination of the ρ-factors involves considerable experimental work.
5. A small uniformly activated rectangular source is required, the construction of which presents difficulties in some cases.

The advantages are:

1. A permanent magnet can be employed to provide a steady magnetic field over long periods. This field can be measured very accurately at leisure, and its use enables different sources to be studied under identical field conditions. This is convenient when direct comparison with standard sources is in question.

2. The resolution can be changed very rapidly by adjusting the widths of the source slit and the counter slit.

3. Comparison of the ρ-factors obtained by use of methods (i) and (ii) of §7 gives an indication of the scattering from the vacuum box walls and baffles present, since scattered electrons will distort the continuous spectrum while leaving the line shape practically unchanged. (This is not strictly an advantage but, if the comparison is satisfactory, disadvantage 3 above is proved to be trivial.) The close agreement, shown in fig. 9, indicates that scattering is unimportant with our arrangement.

A further check on this point is obtained if, after normalization, a good fit results for the overlapping portions of a spectrum obtained at different field settings. (The ratio of the scattered to the continuous spectrum electrons will not be independent of ρ.)

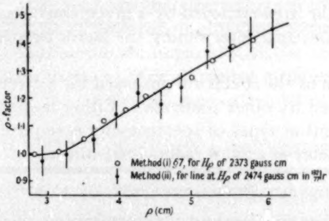


Fig. 9. ρ-factor employed for a given geometry.

4. Line shapes can be studied in detail due to the very small increments in ρ which can be obtained with great accuracy. Again, if agreement is found between the theoretical and experimental line shapes, it can be concluded that the assumed geometrical width of the counter slit is the effective width, and scattering is unimportant.

5. The change over to a photographic method for determining the line positions is very simple.

ACKNOWLEDGMENT

One of us (C. G. C.) wishes to express his appreciation of the opportunities afforded for the latter part of this work by the award of an Imperial Chemical Industries Research Fellowship.

REFERENCES

CAMPBELL, C. G., 1952, *Thesis*, University of Edinburgh.
FEATHER, N., 1940, *Proc. Camb. Phil. Soc.*, **36,** 224.
FEATHER, N., KYLES, J., and PRINGLE, R. W., 1948, *Proc. Phys. Soc.*, **61,** 466.
GURNEY, R. W., 1925, *Proc. Roy. Soc.* A, **109,** 540.
KYLES, J., CAMPBELL, C. G., and HENDERSON, W. J., 1953, *Proc. Phys. Soc.* A, **66,** 519.
LAWSON, J. L., and TYLER, A. W., 1940, *Rev. Sci. Instrum.*, **11,** 6.
LI, K. T., 1937, *Proc. Camb. Phil. Soc.*, **33,** 164.

The Design and Construction of a 1000 kw Water Cooled Solenoid intended for Experiments on Nuclear Alignment and Adiabatic Nuclear Demagnetization

By J. M. DANIELS *
Clarendon Laboratory, Oxford

MS. received 22nd June 1953

Abstract. Details are given of the design, construction and performance of a water-cooled solenoid which develops 50 000 Oe over a volume some 3 in. diameter × 2 in. long, for a continuous input of 1000 kw. The solenoid contains no iron or other ferromagnetic material, and has a disposition of windings to produce the special field distribution required for two-stage adiabetic demagnetization.

§ 1. INTRODUCTION

As part of a long-term programme of research into nuclear paramagnetism at the Clarendon Laboratory, it was decided to attempt experiments on the alignment of nuclei in a large external field, and on adiabatic nuclear demagnetization. Both these projects require that a nuclear paramagnetic should be magnetized to, say, 20% or so of its saturation magnetization at least. Because nuclear magnetic moments are of the order of 10^{-3} times that of the electron, values of H/T are required which are some 10^3 times as large as those needed to magnetize appreciably an electron paramagnetic. Further, the field must be applied continuously, so that the heat of magnetization can be removed. This problem has been discussed at length by various authors, e.g. Simon (1939).

A simple calculation shows that, for a typical case, a field of 70 000 Oe at a temperature of 0·01°K is barely sufficient; and both this field and this temperature are near the limit of present achievement. The heat sink at 10^{-2} deg K to which the heat of nuclear magnetization flows can be produced only by adiabatic demagnetization, and must therefore be situated in a region of small or zero field. The heat of magnetization must therefore be transferred to it by conduction, hence the distance between the nuclear paramagnetic and the heat sink should not be inconveniently large.

It was decided, for reasons not entirely connected with the experiment itself, to produce this field with an air-cored water-cooled solenoid containing no ferromagnetic material. The first problem is, therefore, to find a disposition of windings which will produce a large field in one place; and, in another region some 6 cm long situated some 20 cm or so from the point of maximum field, the field is to be uniformly zero (i.e. H and $\partial H/\partial z$ both zero).

§ 2. FORMULAE FOR THE FIELD PRODUCED BY SOLENOIDS OF VARIOUS SHAPES

Although the calculation of the field and its derivatives at any point on the axis of a solenoid is straightforward, it is nevertheless tedious, and the author knows no publication where these formulae are derived and collected together. Since they are required in the design of the solenoid, they are derived and quoted here.

---

* Now at the University of British Columbia, Vancouver, Canada.

Firstly we shall ignore the fact that a solenoid is wound with discrete wires, and consider that there is a uniform current density flowing in the winding space in the direction of the individual conductors. Take cylindrical polar coordinates (r, θ, z), the z axis being the axis of symmetry of the solenoid, and plane $z=0$ being the plane of symmetry of the solenoid.

Referring to fig. 1, which shows a section through a thick solenoid with rectangular section winding space, we shall often use the following reduced system of measurements:

$$\zeta=\frac{z}{a_1},\qquad \rho=\frac{r}{a_1},\qquad \alpha=\frac{a_2}{a_1},\qquad \beta=\frac{b}{a_1}.$$

a_1 and a_2 are respectively the inner and outer radius of the winding space, and $2b$ is its length.

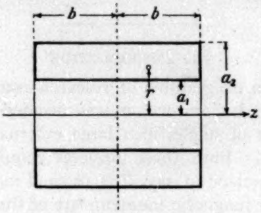


Fig. 1.

The field at a point P $(0, 0, z)$ on the axis of a loop of wire, radius r, carrying a current i e.m.u. is, by Ampère's rule,

$$H=\frac{2\pi r^2 i}{(r^2+z^2)^{3/2}}=\frac{2\pi i}{a_1(1+\zeta^2)^{3/2}}. \qquad \ldots\ldots(1)$$

By symmetry, the direction of H is along the axis.

Integrating this formula with respect to r, we then obtain the following: For an annular coil (i.e. $\beta=0$, $\alpha\neq 0$) (a 'pancake')

$$H=2\pi\sigma\left[\sinh^{-1}\frac{\alpha}{\zeta}-\sinh^{-1}\frac{1}{\zeta}-\frac{\alpha}{(\alpha^2+\zeta^2)^{1/2}}+\frac{1}{(1+\zeta^2)^{1/2}}\right] \qquad \ldots\ldots(2)$$

$$=2\pi\sigma\left[\ln\alpha-\tfrac{3}{4}\zeta^2\left(1-\frac{1}{\alpha^2}\right)+O(\zeta^4)\right] \qquad \ldots\ldots(2a)$$

where σ is the surface current density in e.m.u. cm^{-1} flowing in the plane annulus. From this it follows that

$$\frac{\partial H}{\partial\zeta}=2\pi\sigma\left[\frac{1/\zeta}{(1+\zeta^2)^{3/2}}-\frac{\alpha^3/\zeta}{(\alpha^2+\zeta^2)^{3/2}}\right] \qquad \ldots\ldots(3)$$

and

$$\frac{\partial^2 H}{\partial\zeta^2}=2\pi\sigma\left[\alpha^3\frac{4+\alpha^2/\zeta^2}{(\alpha^2+\zeta^2)^{5/2}}-\frac{4+1/\zeta^2}{(1+\zeta^2)^{5/2}}\right]. \qquad \ldots\ldots(4)$$

For an ordinary single layer solenoid ($\alpha=1$, $\beta\neq 0$) integration with respect to z gives

$$H=2\pi\sigma\left[\frac{\beta-\zeta}{\{1+(\beta-\zeta)^2\}^{1/2}}+\frac{\beta+\zeta}{\{1+(\beta+\zeta)^2\}^{1/2}}\right] \qquad \ldots\ldots(5)$$

whence, for a thick solenoid, by integrating again with respect to r, we get

$$H = 2\pi\tau a_1 \left[\begin{array}{l} (\beta-\zeta)\left\{\sinh^{-1}\dfrac{\alpha}{|\beta-\zeta|} - \sinh^{-1}\dfrac{1}{|\beta-\zeta|}\right\} \\ +(\beta+\zeta)\left\{\sinh^{-1}\dfrac{\alpha}{\beta+\zeta} - \sinh^{-1}\dfrac{1}{\beta+\zeta}\right\} \end{array} \right] \qquad \ldots\ldots(6)$$

where τ is the current density (supposed uniform) in e.m.u. cm^{-2} flowing in the winding space. The expression for H has branch points at $\zeta = \pm\beta$ and different branches of the function are relevant in the different regions. In what follows, we will impose the restriction $\zeta \geqslant 0$ for simplicity.

For $\zeta = \beta$ (i.e. at the branch point),

$$H = 4\pi\tau a_1 \beta \left[\sinh^{-1}\frac{\alpha}{2\beta} - \sinh^{-1}\frac{1}{2\beta} \right]. \qquad \ldots\ldots(6\,a)$$

Differentiating we have

$$\begin{aligned} \frac{\partial H}{\partial \zeta} = 2\pi\tau a_1 \Bigg[& \sinh^{-1}\frac{1}{|\beta-\zeta|} - \sinh^{-1}\frac{\alpha}{|\beta-\zeta|} + \sinh^{-1}\frac{\alpha}{|\beta+\zeta|} - \sinh^{-1}\frac{1}{|\beta+\zeta|} \\ & + \frac{\alpha/(|\beta-\zeta|)}{[1+\{\alpha/(\beta-\zeta)\}^2]^{1/2}} - \frac{1/(|\beta-\zeta|)}{[1+\{1/(\beta-\zeta)\}^2]^{1/2}} \\ & - \frac{\alpha/(\beta+\zeta)}{[1+\{\alpha/(\beta+\zeta)\}^2]^{1/2}} + \frac{1/(\beta+\zeta)}{[1+\{1/(\beta+\zeta)\}^2]^{1/2}} \Bigg]. \qquad \ldots\ldots(7) \end{aligned}$$

Numerical evaluation of this formula is facilitated if use is made of the fact that

$$\frac{\alpha/(|\beta-\zeta|)}{[1+\{\alpha/(\beta-\zeta)\}^2]^{1/2}} = \tanh\left(\sinh^{-1}\frac{\alpha}{|\beta-\zeta|}\right)$$

and similarly for the other three pairs of terms. Differentiating again,

$$\begin{aligned} \frac{\partial^2 H}{\partial \zeta^2} = 2\pi\tau a_1 \Bigg[& \pm \frac{\{1/(\beta-\zeta)\}^4}{[1+\{1/(\beta-\zeta)\}^2]^{3/2}} \mp \frac{\alpha^3/(\beta-\zeta)^4}{[1+\{\alpha/(\beta-\zeta)\}^2]^{3/2}} \\ & + \frac{1/(\beta+\zeta)^4}{[1+\{1/(\beta+\zeta)\}^2]^{3/2}} - \frac{\alpha^3/(\beta+\zeta)^4}{[1+\{\alpha/(\beta+\zeta)\}^2]^{3/2}} \Bigg] \qquad \ldots\ldots(8) \end{aligned}$$

taking the upper signs if $\zeta < \beta$ and the lower signs if $\zeta > \beta$. In particular, if $\zeta = 0$,

$$\left(\frac{\partial^2 H}{\partial \zeta^2}\right)_{\zeta=0} = 4\pi\tau a_1 \left[\frac{1/\beta^4}{\{1+1/\beta^2\}^{3/2}} - \frac{\alpha^3/\beta^4}{\{1+\alpha^2/\beta^2\}^{3/2}} \right]. \qquad \ldots\ldots(9)$$

§3. CALCULATION OF THE DISPOSITION OF WINDINGS FOR THE SOLENOID

Referring to fig. 2 it seemed possible that if an annular coil B were connected in opposition to the main field producing coil A, the field at P could be arranged to have both zero slope and zero curvature. A gapped solenoid C of the type described in a previous publication (Daniels 1950) could be connected to a separate generator and fed with current to neutralize the field at P. The field near P would then have zero slope and zero curvature and hence be reasonably uniform. (Its expansion about P would be as z^3.)

The field H is the sum of the field H_{A} produced by coil A and the field H_{B} produced by coil B. If we arrange that

$$\left(\frac{\partial H_{\mathrm{A}}}{\partial z}\right)_{\mathrm{P}} = \left(\frac{\partial H_{\mathrm{B}}}{\partial z}\right)_{\mathrm{P}} \quad \text{and} \quad \left(\frac{\partial^2 H_{\mathrm{A}}}{\partial z^2}\right)_{\mathrm{P}} = \left(\frac{\partial^2 H_{\mathrm{B}}}{\partial z^2}\right)_{\mathrm{P}},$$

reversing the current in B will then give a resultant field at P of the type desired. Coil A is first designed to produce the field required at O, and to be reasonably efficient as described by Daniels (1950), Cockcroft (1928). If β is made rather less than 2, there is not much loss of efficiency, but a considerable reduction of the stray field at P. If it is not necessary to have a very homogeneous field at O, β can be made as small as 1 with advantage. Next calculate the first and second derivatives of H_A at P, using eqns. (7) and (8), and write $K = (\partial H_A/\partial z)_P$, $L = (\partial^2 H_A/\partial z^2)_P$. Then, for coil B, substituting in eqns. (3) and (4), we have

$$\left.\begin{aligned} \left(\frac{\partial H_B}{\partial z}\right)_P &= \frac{2\pi\sigma}{a_1}\left[\frac{a_1/z}{(1+z^2/a_1^2)^{3/2}} - \frac{\alpha\, a_1/z}{(\alpha^2+z^2/a_1^2)^{3/2}}\right] = K \\ \left(\frac{\partial^2 H_B}{\partial z^2}\right)_P &= \frac{2\pi\sigma}{a_1^2}\left[\frac{1}{\alpha^2}\,\frac{4+\alpha^2 a_1^2/z^2}{(1+z^2/\alpha^2 a_1^2)^{5/2}} - \frac{4+a_1^2/z^2}{(1+z^2/a_1^2)^{5/2}}\right] = L. \end{aligned}\right\} \quad \ldots\ldots(10)$$

These equations must be solved simultaneously, and the variables are σ the surface current density in the annular coil B, a_1 the inside radius of B, z the distance of P from the centre of B, and α.

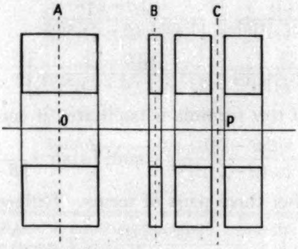


Fig. 2.

Any two of these can be chosen arbitrarily. In general σ depends on the sizes of conductor available, and hence is not continuously variable. It is also limited by the amount of power which it is permissible to dissipate in B, and, in general, should be as small as possible. z, a_1 and α are limited by geometrical considerations.

In order to solve eqns. (10) with a minimum of effort, a certain amount of intelligent guesswork is desirable, and one method is as follows. Put $z = a_1 \sinh\theta = \alpha a_1 \sinh\phi$. Then $\alpha = \sinh\theta/\sinh\phi$, $\theta = \sinh^{-1} z/a_1$, $\phi = \sinh^{-1} z/\alpha a_1$. Equations (10) can then be written:

$$\left.\begin{aligned} \frac{1}{\sinh\theta}\left[\frac{1}{\cosh^3\theta} - \frac{1}{\cosh^3\phi}\right] &= \frac{a_1 K}{2\pi\sigma} \\ \frac{1}{\sinh^2\theta}\left[\frac{4\sinh^2\phi+1}{\cosh^5\phi} - \frac{4\sinh^2\theta+1}{\cosh^5\theta}\right] &= \frac{a_1 L}{2\pi\sigma}. \end{aligned}\right\} \quad \ldots\ldots(11)$$

Eliminate σ; then

$$\frac{\cosh^{-5}\phi - \cosh^{-5}\theta}{\cosh^{-3}\phi - \cosh^{-3}\theta} = \tfrac{1}{3}\left(\frac{a_1 L}{K}\sinh\theta + 4\right) = P, \quad \text{say}$$

or

$$\cosh^{-5}\phi - P\cosh^{-3}\phi + (P\cosh^{-3}\theta - \cosh^{-5}\theta) = 0. \quad \ldots\ldots(12)$$

This equation, which is of degree 5 in $1/\cosh \phi$, can be solved by one of the many numerical methods available. The procedure is, then, as follows:

Fix z and choose some value of a_1 which inspection of eqns. (10) suggests might be reasonable. This gives a value of θ and enables eqn. (12) to be written down with numerical coefficients. Solve eqn. (12) to get the appropriate value of ϕ, from which the corresponding values of α and σ are calculated. The process is repeated for a few different values of a_1 and graphs are drawn of α and σ against a_1. From these graphs a suitable set of values of a_1, α and σ is chosen. If no suitable set of values is obtainable the procedure may be repeated with a different value of z. The field distribution is finally checked by direct calculation using eqns. (2) and (6).

§ 4. CONSTRUCTION

The solenoid was designed according to the method described in the previous section. The mode of construction is very similar to that for the 100 kw water-cooled solenoid described by Daniels (1950), but modifications were necessary to make it more robust.

Coil A consists of six pancake coils wound with $\frac{5}{8}$ in. × 16 s.w.g. copper strip, insulated by wrapping a nylon filament of diameter 0·2 mm round it with a pitch of about 2 mm. The inside radius of the windings is 4·5 cm, the outside radius 11·25 cm, and the length 10·5 cm. The centre of coil B is 14 cm distant from the centre of coil A. For coil B, two copper strips 1 in. × 21 s.w.g., one insulated with nylon of diameter 0·2 mm and the other with nylon of diameter 0·3 mm, are wound in parallel. This is to obtain the correct value of the current density in coil B while it carries all the current from coil A. The inner radius of coil B is 6 cm, and the outer radius 11·44 cm; the total number of turns of each strip is $20\frac{1}{2}$. Coil C consists of four pancake coils of $\frac{5}{8}$ in. × 21 s.w.g. copper strip insulated with nylon of diameter 0·3 mm. The inner radius is 4·5 cm, and the outer radius 11·25 cm. There is a 'uniformizing' gap of 3·08 cm in the centre, and the total length of the windings is 10·8 cm. The centre of coil C is 21 cm from the centre of coil A. The direction of winding of the coils is such that, when all the coils are energized for use, the outside of coil B, and the outside of the pancake of coil C nearest to coil B, are both at earth potential. Coil B is connected permanently in series opposition with coil A. The windings as originally constructed are shown in the photograph, fig. 3.

The pancake coils are wound round central brass rings, as for the 100 kw coil (Daniels 1950), but in this case the rings are $\frac{1}{16}$ in. thick. No trouble has been experienced with these rings. The central brass tube, which runs through the windings and forms part of the water jacket, has 3 in. inside diameter and $\frac{1}{16}$ in. wall thickness. It was coated with polythene by being first warmed, and then wrapped with polythene sheet which melted and adhered. Three or four layers were put on in this way; and after the polythene had cooled it was turned down to the desired diameter.

The external bands which secure the outside of the pancake coils and which form the electrical connections between pairs of pancakes were originally a more robust version of the type used on the 100 kw solenoid, and can be seen in fig. 3. Those on coil C burned out at the screw clip twice on the first two occasions when this coil was tested at high power; we have not yet succeeded in explaining this, but we have replaced them with a new type of outer securing band. This consists of a copper cylinder of about ¼ in. wall thickness turned to the correct size; the pancakes are pressed into it from either end. The inside of the ring is recessed to take the ends of the spokes of the Tufnol spacer. This type of connector has been found to be quite reliable.

The thick spacers filling up large gaps between the various pancakes were fabricated out of Tufnol sheet, as shown in fig. 3. Although these have been quite satisfactory on the whole, those in the central gap of coil C collapsed under the influence of electromagnetic forces while this coil was run at high power. The central portions of the spokes were found to have twisted and sheared off from the ends. These have been replaced by spokes of 'frequentite' (a ceramic), deeply recessed in the outer copper ring, and are now found to be quite satisfactory.

The outer case was cast in 'Hidumidium'—a light alloy of high tensile strength—½ in. thick, and the water manifolds were cast integral with the casing. A thick casing like this is desirable, because the force on the ends due to water pressure at 120 lb in^{-2} can be as much as 8 tons, and the force of repulsion of coils B and C by coil A when carrying the full working current (5000 A) is estimated to be of the order of 5 tons. Adequate water manifolds are needed to cut down unnecessary resistance to the flow of cooling water, and these are most conveniently cast integral with the casing. The casing is cast in four parts—two lids and two cylindrical portions—and each part serves as a terminal. The joints in the casing are insulated with ⅛ in. rubber gaskets, and the bolts holding the various parts together are insulated with Tufnol bushes. We have found that it is desirable that the rubber gaskets should project about ½ in. beyond the outside of the casing, and that the bolt heads should similarly be insulated with large polythene 'skirts', or be totally enclosed. This is a precaution against arcing.

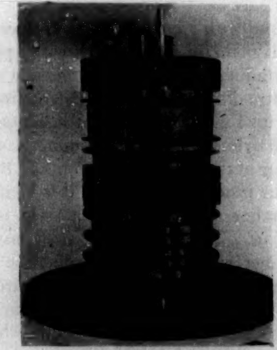


Fig. 3. Windings of the solenoid.

Fig. 4. Exterior view of the solenoid.

The coil was mounted on a wheeled carriage, and moves vertically guided by three pillars and counterbalanced with lead weights. Figure 4 is a view of the completed coil on its carriage.

§ 5. PERFORMANCE

5.1. *Tests at Low Power*

The field at various points on the axis of the solenoid, and its homogeneity, have been measured using search coils and a ballistic galvanometer. In order to measure small differences in field, two identical search coils mounted with adjustable separation on the same holder were connected in opposition in series with a ballistic galvanometer. The field of coil C was found to be quite uniform, as was calculated. The combined field of coils A and B near the centre of coil C, although of the correct magnitude, is not as uniform as was hoped. A graph of field as a function of distance is given in fig. 5. Despite its slope, there is still a region some 5 cm long where the field can be reduced to less than 100 Oe while the solenoid is fully energized, and this is good enough for the experiments envisaged.

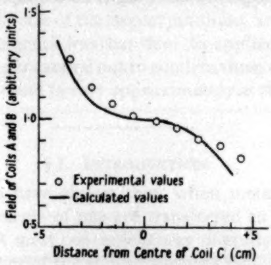


Fig. 5. Observed and calculated values of the field near the centre of coil C due to current flowing in coils A and B.

5.2. *Tests at High Power*

Measurements of field have been made at various places for various currents. The field at the centre of coil C is found to be 14 000 Oe per 1000 A in C; and the field at the centre of coil A is found to be 14 900 Oe per 1000 A in A. Both these are very close to the calculated values, and would give 56 000 Oe and 74 000 Oe for the maximum currents of 4000 A and 5000 A which it is intended shall be supplied to each coil.

The coil is cooled by a forced circulation of water at 200 gallons per minute. The water was specially purified by the Permutit Deminrolit process in order to avoid electrolytic corrosion; its conductivity is about 2 reciprocal megohms per centimetre cube. Although the coil is intended eventually to dissipate 2000 kw continuously, it has been used only for powers up to 1000 kw. This is because the cooling water installation is not yet complete. By simultaneous measurements of current and voltage in the coil during running, one can calculate both the power input and the resistance of the coil. From the resistance of the windings, it is possible to find their average temperature, and hence to obtain information about the heat transfer from the windings to the cooling water. It is found that the greatest temperature drop occurs in the boundary layer between the copper and the water, and that this is of the order of 40°C with an input power of 1000 kw.

ACKNOWLEDGMENTS

I wish to acknowledge my debt to Professor Simon and Dr. Kurti for initiating this work, and for their freely given advice. I am indebted to Dr. E. Ambler for checking the numerical calculations, to Mr. I. I. Boswell of Messrs. Tickford Ltd. and to Mr. J. Milligan of the Clarendon Laboratory for many of the mechanical details, and to Mr. F. N. H. Robinson for assistance in testing the solenoid.

This work was carried out partly during the tenure of a Department of Scientific and Industrial Research maintenance grant, and partly during the tenure of a Nuffield Foundation Research Fellowship.

REFERENCES

COCKCROFT, J. D., 1928, *Phil. Trans. Roy. Soc.* A, **227,** 317.
DANIELS, J. M., 1950, *Proc. Phys. Soc.* B, **63,** 1028.
SIMON, F. E., 1939, *Comptes Rendus Congrès sur le Magnétisme, Strasbourg,* **3,** 1.

A Quantitative Study of the Wear Process

By E. RABINOWICZ

Lubrication Laboratory, Department of Mechanical Engineering,
Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.

MS. received 24th June 1953

Abstract. In this paper a discussion is presented of the determination of the mass of wear fragments using autoradiographic techniques. The size distribution of copper wear fragments transferred during sliding on mild steel is measured experimentally, and using these data the size distribution of the metallic junctions is calculated. An estimate is presented of the effect on the junctions of an increase of load, and it is shown that the size of the largest junctions, and hence the mass of the largest wear particles, varies much less than does the applied load. In experiments of mild steel sliding on copper carried out to confirm these calculations, the mass of the largest fragments was found to vary approximately as the 0·3 power of the load.

---

§ 1. INTRODUCTION

MANY experiments have shown that, when metal surfaces are slid over each other, fragments of one are transferred to the other (Bowden and Tabor 1950). A most convenient way of studying the phenomenon in detail consists of making one of the sliding surfaces radioactive and, after sliding, studying the transfer of radioactivity to the other by means of a photographic film placed on the surface (Rabinowicz and Tabor 1951). Each fragment then gives rise to a circular blackened region on the photographic film (fig. 1, Plate). This article will describe in detail the measurement of the size of the fragments and discuss the data obtained in some typical sliding experiments in the light of recent theories of the wear process.

§ 2. MEASUREMENT OF MASS OF FRAGMENTS

To a first approximation the radioactive fragment situated near the photographic film may be regarded as a point source touching the film which is assumed to have zero thickness. In this case the radioactive intensity I at any point on the film will be proportional to the mass m of the fragment and inversely proportional to the square of its distance x from the fragment.

Thus
$$I = km/x^2 \qquad \ldots\ldots(1)$$
where k is a constant.

In practice a number of factors are present which produce complications, namely finite size of fragment, finite distance of fragment from the film, and finite thickness of the film. These will produce a geometric effect and also scattering and shielding of the radiation. Using the β- and γ-rays of this investigation the latter effects are small and have been neglected. The geometrical correction can be stated with sufficient accuracy in terms of a parameter t which is essentially the distance from the centre of gravity of the fragment to the centre of the film.

Then $$I = \frac{km}{t^2 + x^2}. \qquad \ldots\ldots(2)$$

In our experiments, t was a distance of the order of several times 10^{-3} cm.

Figure 2 represents an autoradiograph of three fragments of different size. Since the response curve of a photographic film is essentially logarithmic, the intensities have also been plotted logarithmically. It will be seen that the centre of each of the larger circles is black, there is an intermediate grey region in which the intensity gradually falls off, and eventually a white region in which the radioactive blackening is little greater than the background fogging. In the case of the smaller fragments the centre black region is absent.

In taking quantitative measurements, it is convenient to produce the images of high contrast so as to make the grey intermediate region as small as possible. When this is done, it becomes possible to measure the diameter d of each region, namely the circle inside which the blackening exceeds some definite arbitrarily chosen shade of grey, with fair precision. Figure 2 shows that in the case of the larger images, the diameters d_1 and d_2 so measured are very nearly as great as would be given by the ideal formula of eqn. (1) (shown as the broken curve), but in the case of the smaller image the measured diameter d_3 is far too small.

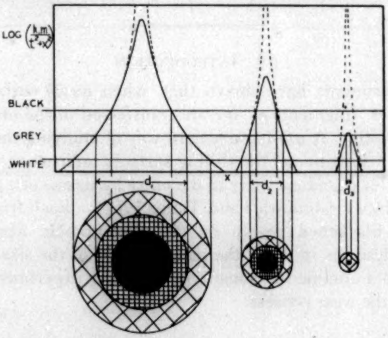


Fig. 2. Autoradiographic appearance of three fragments of different size together with the actual (full line) and idealized (broken line) intensity of blackening.

Since the intensities at the outside of the black image are equal, we have for the larger fragments

$$I = 4km_1/d_1^2, \qquad I = 4km_2/d_2^2, \quad \text{therefore} \quad m_1/m_2 = d_1^2/d_2^2. \quad \ldots\ldots(3)$$

In our experiments an enlarged negative was taken of the original autoradiograph, and, further enlarged, highly contrasty positives were obtained. By varying the exposure times during printing the size of the various circles could readily be varied, and all fragments were measured under conditions in which deviations from the inverse square law were not serious. The limit to the measurement of still smaller fragments was placed by the stray background and by the presence of radiation from nearby large fragments.

Since only the ratio of the masses of the fragments is obtained by this procedure, the measurements of radioactive intensity are only relative. Two separate methods were used to obtain the actual masses of the larger fragments. In the first the blackening produced by the fragment was compared with that produced by a thin foil of known thickness of the same metal. By simple geometry it can easily be shown that if the intensity at a radius x_1 from the centre of the fragment is equal to the (uniform) intensity produced by the foil under the same conditions, then

$$m_1 = 2\pi\rho x_1^2 h \qquad \ldots\ldots(4)$$

where h is the thickness of the foil and ρ the density of the metal.

Alternatively the length and width of the fragment were measured under the microscope, the height of the fragment was taken to be equal to half the width, and the volume calculated assuming the fragment to be half an ellipsoid of revolution. In the present case these two methods gave values differing by 30% and an average was adopted. Previous measurements (Rabinowicz 1951) had shown that over a wide range of size the mass of the fragments was proportional to the 3/2 power of their projected area, and this result, suggesting that the wear mechanism is the same for all fragments irrespective of size, has determined the choice of the wear model to be considered later.

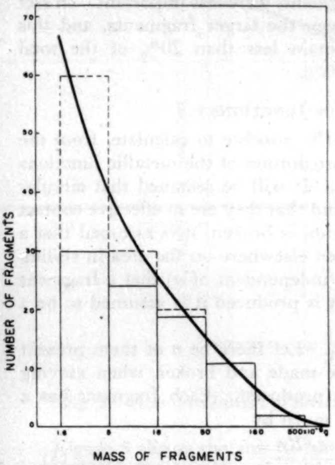


Fig. 3. The size distribution of copper fragments. The full line shows the measured number of fragments, the broken line takes into account the number hidden by the presence of nearby fragments.

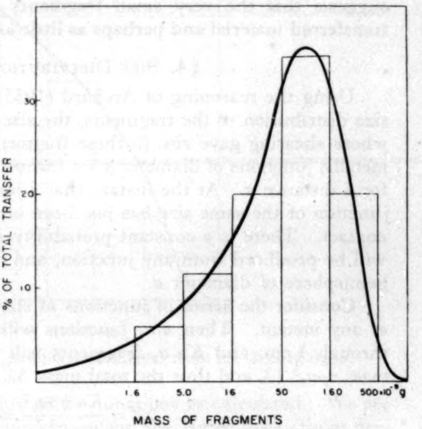


Fig. 4. Plot of the relative importance of the size-groups of wear particles shown in fig. 3 as a percentage of the total mass of wear.

§ 3. Size Distribution of Fragments

Sliding experiments were carried out using clean, radioactive copper containing some of the isotope ^{64}Cu on mild steel at a load of 2000 g and a speed of 0·01 cm sec^{-1}, and the size of the fragments was measured along 1 cm of the track. The data obtained from the black circles of fig. 1 and similar photographs are shown in fig. 3 in which the number of fragments is plotted as a function of their mass. Each column covers a mass range of a factor of $10^{1/2}$. In the case of the lighter fragments, complications were caused by the fact that only those fragments that were a considerable distance from the nearest large fragment could be observed. However, a correction was applied by assuming a random distribution of the fragments and estimating the number shielded by the larger fragments.

It will be seen that the smaller fragments are by far the most numerous. However, most of the mass of transferred material is actually contained in the larger fragments. Figure 4, obtained by multiplying each column in fig. 3 by its average mass, shows the percentage of the total transfer contained in each column, it being assumed that the end of the curve corresponding to those fragments too small to be observed 'tails off' smoothly. That these very small fragments do in fact contribute little to the total wear is seen by examination of the original autoradiograph. If the track consisted of a dense continuous background with here and there circular regions of greater blackness, we would assume that the fragments too small to be resolved contained most of the mass of transferred material and that the bigger fragments were less important. In fact the track is quite clear in the regions between the larger fragments, and this suggests that the very small fragments contain less than 20% of the total transferred material and perhaps as little as 10%.

§4. SIZE DISTRIBUTION OF JUNCTIONS

Using the reasoning of Archard (1953), it is possible to calculate, from the size distribution of the fragments, the size distribution of the metallic junctions whose shearing gave rise to these fragments. It will be assumed that circular metallic junctions of diameter a are formed, and that they are in effective contact for a distance a. At the instant that a junction is broken, it is assumed that a junction of the same size has just been created elsewhere on the area in sliding contact. There is a constant probability K (independent of a) that a fragment will be produced from any junction, and if it is produced it is assumed to be a hemisphere of diameter a.

Consider the series of junctions of size a_1. Let there be n of them present at any instant. Then n/a_1 junctions will be made and broken when moving through 1 cm, and Kn/a_1 fragments will be produced. Each fragment has a mass $\pi\rho a_1^3/12$, and thus the total mass ΔZ is given by

$$\Delta Z = \frac{Kn}{a_1}\frac{\pi\rho a_1^3}{12} = \frac{\pi\rho a_1^2 Kn}{12}. \qquad \ldots\ldots(5)$$

Now at any instant the total area ΔA due to this series of junctions will be $\pi n a_1^2/4$, and this will support a load ΔL where

$$\Delta L = p\Delta A = \pi n a_1^2 p/4 \qquad \ldots\ldots(6)$$

p being the flow pressure of the softer metal as used in the well-known plastic flow condition $L = pA$.

Hence
$$\frac{\Delta Z}{\Delta L} = \frac{\rho K}{3p} \qquad \ldots\ldots(7)$$

and summing for all sizes of junctions we get

$$\frac{Z}{L} = \frac{\rho K}{3p}. \qquad \ldots\ldots(8)$$

To derive from the size distribution of fragments a size distribution of junctions, we use the formula

$$\frac{\text{No. of junctions at any instant}}{\text{No. of fragments on sliding 1 cm}} = \frac{n}{nK/a} = \frac{a}{K}. \quad \ldots\ldots(9)$$

K can readily be calculated from eqn. (8). One modification should be made to the reasoning of Archard, who took p as the bulk hardness of the metal. There is evidence (Moore 1948, Rabinowicz 1953) that even for surfaces in localized contact there is a considerable difference between the real and the apparent areas of contact, the latter being greater by a factor of about 3, due to localized work-hardening at the asperities and to the increased strength of small regions of a metal. Hence we shall take the real value of p for the copper specimen as being three times the value given by the Vickers hardness test, namely 285 rather than $95 \times 10^5\,\text{g cm}^{-2}$.

Using the data $L = 2 \times 10^3\,\text{g}$, $Z = 10^{-6}\,\text{g cm}^{-1}$, $\rho = 9$, we get $K = 5 \times 10^{-3}$.

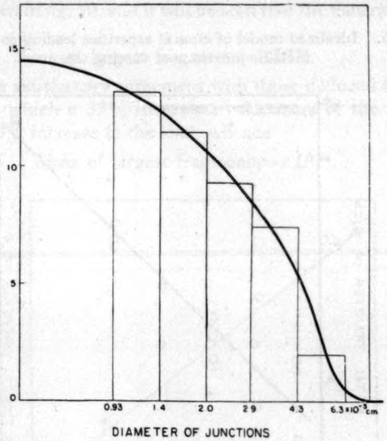


Fig. 5. Size distribution of metallic junctions calculated from the data of fig. 3.

Figure 5 shows the size distribution of the junctions as calculated. We see that the smaller junctions are more numerous than the larger ones; there may be a maximum at about 4×10^{-4} cm.

The graph of percentage of total area in each column as a function of the area of the junctions, obtained by multiplying each column in fig. 5 by a^2, has the same shape as fig. 4 (both consisting essentially of fig. 3 multiplied by a^3) and is, therefore, not shown separately.

§ 5. Effect of an Increase in Load

5.1. *Theoretical*

It is of interest to consider the effect of an increase of load on the size distribution shown in fig. 5. We shall take a simple model in which the harder metal surface is perfectly plane, and the softer surface has asperities that are circular cones of constant angle but different height (fig. 6). When the load is increased the surfaces are pressed closer together, plastic flow of the softer metal takes place, and the diameter of each junction is increased by the same amount. This will increase the relative importance of the smaller junctions, since their relative increase in area will be greater. At the same time, owing to their increase in size many junctions will fall into new size-brackets. The effect of these changes on the size distribution of the junctions has been calculated for an increase in load sufficient to increase the diameter of the largest junctions by 10%. Although the area of these largest junctions is increased by only 21%, the total area in contact, and hence the load that can be carried, is increased by as much as 90%. Apart from this the shape of the curve is nearly the same as at the lower load.

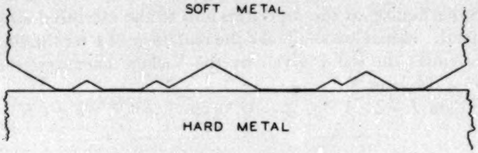


Fig. 6. Idealized model of conical asperities leading to circular metallic junctions of varying size.

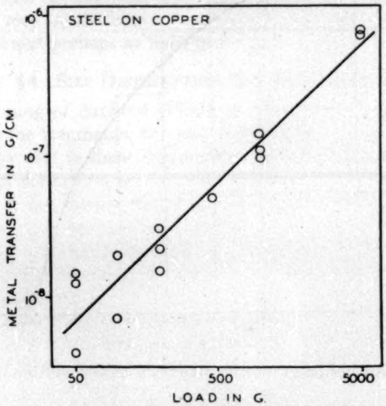


Fig. 7. Mass of steel fragments transferred on to a clean copper surface plotted on a logarithmic scale as a function of the applied load. The line has a slope of 1.

5.2. *Experimental*

The increase in fragment size with load was studied in a separate experiment. In this case the sliding arrangement was rather different, being of a radioactive mild steel hemisphere containing ^{55}Fe on a copper surface, at a speed of

0·02 cm sec^{-1}. The transfer of steel to the copper was measured. Figure 7 shows a plot of mass of metal transferred as a function of load on a logarithmic scale, and it will be seen that a line of slope 1 fits the points well, suggesting that in this case Z was proportional to L. Experiments such as this, carried out over large ranges of load, lend real support to the concept of a constant coefficient of wear K. For this experiment K had a value of $1{\cdot}0 \times 10^{-3}$.

Attempts were made to measure the sizes of the fragments at the various loads. Figures 8 and 9 (Plate) show the tracks at 50 and at 5000 g. Owing to the very different density of fragments in the two cases the simplest and perhaps most satisfactory procedure is to compare the largest fragments on the two tracks. Making allowances for the presence of nearby fragments at the 5000 g load which raises the background and consequently makes each fragment appear bigger, we may estimate that the largest fragments in fig. 9 are some 5 times larger than those in fig. 8, although owing to the great difference in appearance of the two photographs it is not easy to make a precise estimate. The sizes of the largest fragments at other loads are shown in fig. 10, and it will be seen that the data can be represented by the equation

$$\text{Mass of largest fragments} = c\,L^{0{\cdot}29}.$$

These results are in satisfactory agreement with those deduced from the conical-asperity model, in which a 33% increase in the mass of the largest fragment corresponds to a 90% increase in the load, whence

$$\text{Mass of largest fragments} = c\,L^{0{\cdot}44}.$$

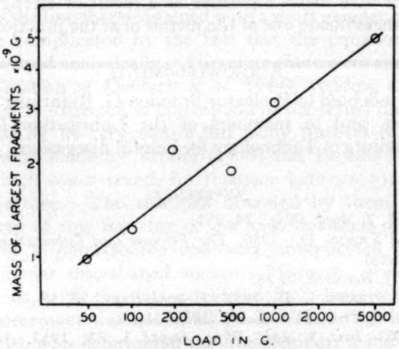


Fig. 10. Plot of mass of largest steel fragments on a copper surface as a function of load. The straight line has a slope of 0·29.

§ 6. Discussion and Summary

These experiments supply further detailed information about the nature of the sliding process. For one representative friction run, measurements were made of the size distribution of fragments, and it was found that the smaller fragments are more numerous while the larger fragments account for most of the mass of wear.

Making plausible assumptions as to the mechanism of wear, the frequency distribution of the metallic junctions was then calculated. Again it was found that the smaller junctions are more numerous (although the disparity is less than in the case of the fragments); however, the larger junctions account for most of the contact area.

Using a simple model, the effect of an increase in load on the frequency distribution of the junctions was then calculated. A 21% increase in the area of the larger junctions leads to a 90% increase in the total contact area; otherwise, the final frequency distribution is very similar to the initial one. In an experimental test a hundredfold increase in load produced a fivefold increase in the mass of the largest transferred fragments (and hence a 2·9-fold increase in the area of the largest junctions), and both the theoretical and experimental data confirm previous deductions (Rabinowicz and Tabor 1951, Rabinowicz 1953, Burwell 1953) that an increase in load acts mainly in increasing the number of junctions and fragments and only to a secondary extent in increasing their size.

These observations of the effect of load on the friction and wear process do not, however, apply universally. In many cases an increase in load radically changes the sliding conditions; to give one example, if the transferred wear fragments are able to interact with each other, then a small increase of load may lead to a very large increase in the average mass of the fragments and in the total mass of transfer (Rabinowicz 1953). Other mechanisms leading to other than proportionate change of wear with load have been suggested (Burwell and Strang 1952). However, the simple wear relationship characterized by a constant, non-dimensional wear coefficient K (analogous to the friction coefficient μ) appears to correspond to the basic wear mechanism, i.e. removal of individual fragments during the encounter of individual metal asperities as a result of strength at the metallic junction and weakness inside one of the metals near the junction.

ACKNOWLEDGMENTS

The author is indebted to Professor Brandon G. Rightmire for his encouragement and interest, and to members of the Lubrication Laboratory at the Massachusetts Institute of Technology for helpful discussions.

REFERENCES

ARCHARD, J. F., 1953, *J. Appl. Phys.*, **24,** 981.

BOWDEN, F. P., and TABOR, D., 1950, *The Friction and Lubrication of Solids* (Oxford: Clarendon Press).

BURWELL, J. T., 1953, *Research*, **6,** 255.

BURWELL, J. T., and STRANG, C. D., 1952, *J. Appl. Phys.*, **23,** 18.

MOORE, A. J. W., 1948, *Proc. Roy. Soc.* A, **195,** 231.

RABINOWICZ, E., 1951, *Brit. J. Appl. Phys.*, *Suppl.* 1, 82; 1953, *Amer. Soc. Lub. Eng.*, in the press.

RABINOWICZ, E., and TABOR, D., 1951, *Proc. Roy. Soc.* A, **208,** 455.

Investigations on a Short Magnetic Lens Spectrometer

By M. K. BANERJEE and A. K. SAHA
Institute of Nuclear Physics, Calcutta, India

Communicated by N. Feather; MS. received 25th March 1953, and in final form 10th August 1953

Abstract. In the course of an attempt to find reversible relations between the parameters describing the trajectory of an electron in a short magnetic lens spectrometer, it was found that the magnetic lens can be surprisingly well represented by a thick converging type optical lens which satisfies Newton's laws for the conjugate planes and the tangent relation far outside the paraxial range where the lens power calculated by the paraxial method can be in error by as much as 50%. Expressions have been obtained for the focal length as a function of the object angle and momentum of the electron. Various other properties of the lens spectrometer have been investigated. In particular, the use of a continuous baffle system to improve the resolution and line-shape for an extended source has been studied by an easy and accurate method.

§1. INTRODUCTION

In determining the line shape obtained with a short magnetic lens spectrometer, the matter is complicated by the fact that the equation of the trajectory could not be solved analytically. Therefore there were approximate methods (paraxial approximation of Deutsch *et al.* 1944), yielding a line-shape, which however could not be expected to correspond to the actual line-shape. Furthermore, the method by its very nature hid many features of the spectrometer. Later, attempts were made by Verster (1949) and Perkins and Solbrig (1951), the essence of which was a search for relations between parameters describing the electron trajectory. The relations obtained by them revealed the very important property of ring-focusing of the lens, but their method of obtaining the line-shape for a point-source was very complicated and could not be extended to a circular disc-shaped source. Moreover, we tried to test their parametric relations, with the data calculated for a lens we are designing, and found that the agreement was not at all good. There has been an attempt by Keller *et al.* (1950), who determined the line-shape by a method which appears to us to be rather approximate. In this paper we have used the method of parametric relations to obtain information about the lens. The relations that we have obtained between the different parameters are remarkable in the sense that they have a close resemblance to the relations for a thick optical lens. Moreover, the determination of the line-shape for a point-source becomes very simple and its extension to a disc-source line-shape could be carried out at once, because our parametric relations were at every stage completely reversible. The parametric relations showed clearly how the resolutions could be considerably improved by the use of a continuous baffle system. This point was implied in the work of Perkins and Solbrig, but they apparently overlooked it.

§2. Description of the Instrument

The constructional design of the instrument is almost similar to that of Deutsch *et al.* (1944). The principal features of the instruments, as far as they concern us here, are shown in fig. 1. The magnetic field is produced by an air-core coil. Besides axial symmetry, there is a plane of symmetry normal to the axis of symmetry at the centre of the length of the coil. This plane will be referred to as the 'central plane'.

The expressions for the field and the magnetic vector potential are given by Deutsch *et al.* (1944). $H_0(z)$, the field along the axis at a point on the axis, and its first two derivatives are shown in fig. 2.

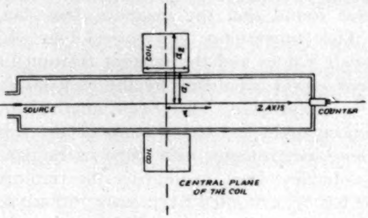


FIG 1

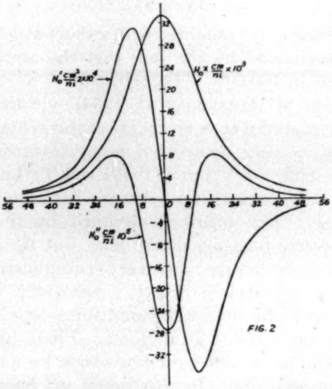


FIG. 2

§3. Calculation of Trajectories

Keller *et al.* (1950) have given a very convenient form for the equation of the trajectory for rays of zero angular momentum which is

$$r''\frac{K^2-A^2}{1+(r')^2}+A\frac{\partial A}{\partial r}-r'A\frac{\partial A}{\partial z}=0 \qquad \ldots\ldots(1)$$

where A is the magnetic vector potential, $r'=\partial r/\partial z$, $r''=\partial^2 r/\partial z^2$, and K is the momentum in units of $300\,e\mathrm{v}/\mathbf{c}=\boldsymbol{m}v\mathbf{c}/\boldsymbol{e}$.

In the paraxial case, i.e. when the rays are close to the axis, the orbit equation reduces to

$$\frac{d^2r}{dz^2} = -\frac{r}{4K^2}H_z^2. \qquad \ldots\ldots(1\,a)$$

It is easier to handle this equation rather than eqn. (1) by numerical methods. For the special case $H_z = H_0/(1+z^2/a^2)$, numerical integration of (1 *a*) is not necessary and Glaser has obtained expressions for focal length measured from the mid-plane and principal plane in analytic forms. Paraxial approximation is very useful in dealing with the electron optical problem of electron microscopes. There the angular aperture is of the order of 10^{-2} radian. The small defects in image due to various aberrations can also be quantitatively determined by a method due to Glaser. But paraxial approximation cannot be recommended for treating problems of the lens spectrometer even with subsequent corrections for aberration, because the minimum object angle in this case is over 5°. To illustrate the error involved in paraxial approximation we have calculated two trajectories for $K = 0{\cdot}4ni$ using (1 *a*) and have compared the values of V and

	V		$\tan\beta$	
$\tan\alpha$	paraxial	rigorous	paraxial	rigorous
0·120 00	79·86 cm	67·68 cm	0·076 33	0·089 52
0·190 00	79·81 cm	52·51 cm	0·120 97	0·181 25

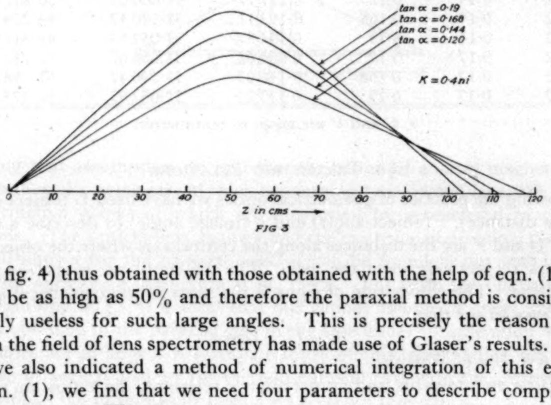


FIG 3

tan β (see fig. 4) thus obtained with those obtained with the help of eqn. (1). The error can be as high as 50% and therefore the paraxial method is considered to be entirely useless for such large angles. This is precisely the reason why no worker in the field of lens spectrometry has made use of Glaser's results. Keller *et al.* have also indicated a method of numerical integration of this equation. From eqn. (1), we find that we need four parameters to describe completely an electron trajectory. These are:

(*a*) $D = \dfrac{ni}{K} = \dfrac{\text{Ampere turns of the coil}}{\text{Momentum in units of } 300\,\mathbf{ev/c}}.$

(*b*) α = angle of emission of the electron from the source.

(*c*) s = perpendicular distance of the source point from the axis of symmetry.

(*d*) Coordinate of the source point measured along the z axis from a suitable origin. The source is a plane surface mounted normal to the z axis at a distance of 50 cm from the central plane. This coordinate is therefore not a variable and is denoted by U_0 (= 50 cm).

Combining four different values of D and $\tan\alpha$, sixteen trajectories have been calculated for $s=0$. For $s=0{\cdot}2$ cm six trajectories have been calculated using a combination of two different values of D and three different values of $\tan\alpha$.

Trajectories for $D=2{\cdot}5$ and $s=0$ are shown in fig. 3. It is to be noted that beyond a distance of about 30 cm from the central plane the trajectories are practically straight. The data of the calculation are given in the table.

s	$(K/\kappa i)^2$	$\tan\alpha$	$\tan\beta$	U	V
0	0·14	0·12	0·114 13	50·000 00	52·798 73
0	0·14	0·144	0·144 67	50·000 00	49·812 67
0	0·14	0·168	0·181 95	50·000 00	46·048 25
0	0·14	0·19	0·224 20	50·000 00	42·034 79
0	0·15	0·12	0·100 82	50·000 00	59·962 30
0	0·15	0·144	0·128 63	50·000 00	56·205 55
0	0·15	0·168	0·162 70	50·000 00	51·710 87
0	0·15	0·19	0·201 55	50·000 00	46·929 59
0	0·16	0·12	0·089 52	50·000 00	67·685 43
0	0·16	0·144	0·115 37	50·000 00	62·930 74
0	0·16	0·168	0·146 53	50·000 00	57·644 16
0	0·16	0·19	0·181 25	50·000 00	52·512 00
0	0·17	0·12	0·079 30	50·000 00	76·679 44
0	0·17	0·144	0·102 69	50·000 00	70·898 43
0	0·17	0·168	0·131 07	50·000 00	64·630 50
0	0·17	0·19	0·163 05	50·000 00	58·563 63
0·2	0·14	0·12	0·121 77	51·666 67	50·837 70
0·2	0·14	0·168	0·193 11	51·190 47	44·224 31
0·2	0·14	0·19	0·234 87	51·052 63	40·812 93
0·2	0·17	0·12	0·086 68	51·666 67	72·283 34
0·2	0·17	0·168	0·140 03	51·190 47	61·798 47
0·2	0·17	0·19	0·173 25	51·052 63	56·175 41

s, U and V are given in centimetres.

§4. PARAMETRIC RELATIONS

Following the practice of geometrical optics we have used U (object distance), V (image distance), α (object angle) and β (image angle) to describe a trajectory (fig. 4). U and V are the distances along the central axis where the object and the image rays respectively meet the axis. The distances are measured from the central plane. The convention of signs is the same as in geometrical optics.

It is easy to see that

$$U=U_0+s\cot\alpha. \qquad \ldots\ldots(2)$$

So for $s=0$, $U=U_0=50$ cm.

As a first step towards finding parametric relations, the ratios $\tan\alpha/\tan\beta$ for the sixteen trajectories with $s=0$ were plotted against V and a linear relation was found of the form

$$\frac{\tan\alpha}{\tan\beta}=a+bV. \qquad \ldots\ldots(3\,a)$$

b was found to be independent of α and D. Moreover, it was found that the quantity a is numerically equal to $1-bU_0$. Proceeding on this clue, we write

$$\frac{\tan\alpha}{\tan\beta}=1+b(V-U_0), \qquad \ldots\ldots(3\,b)$$

but even in this form the relation was not reversible. Then it was argued that instead of assuming that the principal planes of the lens coincide with the central plane, attempts should be made to locate them. Anticipating an analogy with the converging type thick optical lens, the principal planes were taken to be crossed. Symmetry about the central plane ensures that the two principal planes will be equidistant from the central plane (fig. 4).

Let the distances measured from the principal planes be denoted by small letters. If δ is the distance between the central plane and a principal plane, we have

$$u = U + \delta = U_0 + \delta + s \cot\alpha = u_0 + s \cot\alpha \qquad \ldots\ldots(4\,a)$$

$$v = V + \delta. \qquad \ldots\ldots(4\,b)$$

δ can be determined from the property of the principal planes, viz.

$$\frac{\tan\alpha}{\tan\beta} = \frac{v}{u_0} = \frac{V+\delta}{U_0+\delta} = 1 + \frac{V-U_0}{U_0+\delta}. \qquad \ldots\ldots(5)$$

(3 b) and (5) are of the same form if δ is taken to be independent of D and α. This surprising conclusion will be discussed again at the end of this section.

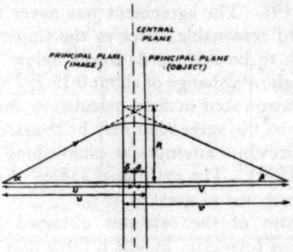


FIG. 4

Making that assumption we find that the action of a short magnetic lens is similar to that of a symmetrical thick converging type optical lens, the locations of the principal planes of which are unaffected by geometrical or chromatic aberrations. The optical lens analogy and the existence of principal planes were quite well known for the paraxial case, as can be found in any standard book on electron optics (e.g. Zworykin *et al.* 1945). But from the discussion given in the beginning of § 3, it is clear that one cannot assume, ***a priori***, the validity of results obtained in the paraxial case in the case under consideration where the object angles are so large. A fortunate result of this investigation has been the restoration of optical analogy for the non-paraxial case with the aforementioned modification. We have, in our case, $\delta = 1/b - U_0 = 1{\cdot}98$ cm.

As a next step, we assume the validity of the formula

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}. \qquad \ldots\ldots(6)$$

f was calculated for the sixteen trajectories with $s = 0$ and $u = u_0$ as a consequence. It showed a marked dependence on D and α. But instead of employing α as a variable we introduce a quantity

$$R = u \tan\alpha = v \tan\beta \qquad \ldots\ldots(7)$$

to express the geometrical aberration. Being a quantity common to the object and the image spaces, its use keeps the equation in a reversible form.

It was found that f could be expressed as

$$\frac{1}{f}=\mu(R)+\nu(R)D^2 \qquad \ldots\ldots(8)$$

$$\mu(R)=\mu_0+\mu_1R^2+\mu_2R^4; \quad \nu(R)=\nu_0+\nu_1R^2+\nu_2R^4 \qquad \ldots\ldots(9)$$

where

$\mu_0=0{\cdot}01663\ \mathrm{cm}^{-1}$; $\mu_1=1{\cdot}54663\times10^{-4}\ \mathrm{cm}^{-3}$; $\mu_2=-1{\cdot}05013\times10^{-6}\ \mathrm{cm}^{-5}$,
$\nu_0=0{\cdot}00485047\ \mathrm{cm}^{-1}$; $\nu_1=-1{\cdot}97731\times10^{-5}\ \mathrm{cm}^{-3}$; $\nu_2=2{\cdot}11733\times10^{-7}\ \mathrm{cm}^{-5}$.

Equations (5), (6), (8) and (9) were obtained from a set of data where u was constant. In order to test the relations, in particular the correctness of introduction of u as a variable in the form (5) and (6), for different values of u, six trajectories were calculated having $s=0{\cdot}2$ cm. For these, we have

$$u=u_0+s\cot\alpha=[51{\cdot}98+0{\cdot}2/\tan\alpha]\ \mathrm{cm}. \qquad \ldots\ldots(10)$$

The data obtained from these calculated trajectories were compared with those from (5), (6), (8) and (9). The agreement was never worse than 0·4%. This error can be considered reasonable, since in the trajectory calculation we have assumed the curvature to be zero at distances below 16 cm and over 86 cm. Actually the slope will show a change of about 0·15–0·2% on either side. Besides some error has been introduced in field calculation and numerical integration. Considering these points, the agreement may be considered to be excellent.

There have been previous attempts at establishing parametric relations by Perkins and Solbrig (1951). The calculated values of v and β did not fit their relations. The departure was as great as 15%.

The happiest feature of the relations obtained above is the complete analogy with optical lens formulae, in which form they are perfectly reversible. Reversibility is a necessary physical condition which any general parametric relation will have to satisfy for a system, whose Hamiltonian is unaltered on reflection of z coordinate.

Equation (9) shows that spherical aberration of fifth order is also present in a lens spectrometer. In the electron microscope where the lens aperture is extremely small, consideration of cubic aberration is enough. But here α is as large as 10·5°, so the higher order aberration must be considered. The first term in (8) expresses the pure geometrical aberration, while the second expresses the combined chromatic and spherical aberrations.

It will be difficult to define a spherical aberration constant for a lens as has been done by Verster (1949) or Butt (1949). But as the spherical aberration constant is neither very useful nor so fundamental, the use of the proposed relations is not handicapped by the aforesaid defect.

The reason why δ appears to be constant is not very clear. The order of accuracy of our calculation, which is within 0·4% for v etc., is only 10% for δ, so a small variation in δ, if any, will not be apparent at all in our formulation. It will be an impossible task to obtain any knowledge about δ by adopting the method of combination of lenses. A magnetic lens is composed of a continuity of infinitesimally thin lenses, each of which is bounded by two close members of a family of surfaces given by the equation

$$\text{Kinetic Energy}+\frac{\mathbf{e}}{\mathbf{c}}(\mathbf{A}\,.\,v)=\text{constant}.$$

Besides having a continuously changing refractive index, the surfaces of the elementary lenses are not spherical. These facts make the consideration extremely complicated.

§ 5. RING FOCUSING FOR A POINT SOURCE ON THE AXIS

In fig. 3 it can be seen that due to high spherical aberration, the image of a point source on the axis is not another point on the axis. In fact, there is no point-focusing at all. The nearest approach to it is an annular zone of least confusion, which is commonly described as 'ring focusing'. It can be easily seen that the ring occurs where the extreme rays intersect.

The coordinates (z_0, r_1) of the intersection of the extreme rays give the location of the ring plane and its inner radius. Rays emitted with intermediate angles will have $r(z_0)$ greater than r_1. The maximum of these values, r_2, gives the outer radius of the ring.

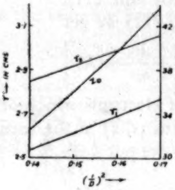


FIG 5a

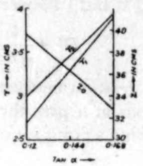


FIG 5b

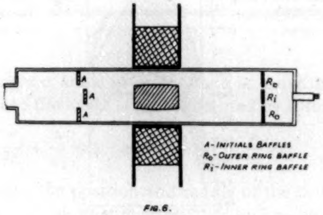


FIG. 6.

As long as the ring is located in the weak field region z_0, and r_1 can be calculated be describing the linear portion of the trajectory by the equation:

$$\frac{r}{R} + \frac{z}{v} = 1. \qquad \ldots\ldots(11)$$

Choosing the extreme values of $R = u \tan\alpha$ and the value of D for which the ring is to be found, ray intersection can be calculated from the equations:

$$r_1 = (v_2 - v_1) \bigg/ \left(\frac{v_2}{R_2} - \frac{v_1}{R_1}\right) \qquad \ldots\ldots(12\,a)$$

$$z_0 = (R_2 - R_1) \bigg/ \left(\frac{R_2}{v_2} - \frac{R_1}{v_1}\right) \qquad \ldots\ldots(12\,b)$$

where v_1 and v_2 are the values of image distance for $R = R_1$ and R_2, the two extreme values of R. A method of calculating r_2 is described in § 7.

z_0, r_1 and r_2 depend on the limits on angle of emission and D. The dependence has been shown in figs. 5 (*a*) and 5 (*b*). The comparatively sharp focal property of the ring has been utilized to obtain improved resolution. A baffle with an annular opening, which allows the extreme rays, forming the ring, to pass through, is placed at the location of the ring. Such a baffle system is shown in fig. 6. These will be referred to as 'ring baffles'.

The value of D for which the ring has been calculated is denoted by D_0 and the corresponding momentum is sometimes called the focusing momentum. The ring baffles transmit all the electrons that are allowed by the initial baffle system (fig. 6) limiting R_1 and R_2. For a slightly different momentum, the ring is displaced and the baffle system will transmit only a part of the focused electrons. This is how we get a line-shape whose peak occurs at $D = D_0$.

§6. Calculation of Line-Shape

Let us consider a point (r, z) in a region in the image space where the field is small. A ray through the point is given by eqn. (11).

Eliminating v and f between (6), (8) and (11) we get

$$D^2 = \frac{1}{\nu(R)}\left[\frac{1}{z}\left(1-\frac{r}{R}\right)+\frac{1}{u}-\mu(R)\right]. \qquad \ldots\ldots(13)$$

This gives the relation between D and R of electrons which originating from a point on the axis at u pass through the point (r, z) in the image space. Since we are considering a point source on the axis we put $u = u_0 = 51{\cdot}98$ cm. In fig. 7 eqn. (13) has been represented in a curve whose abscissa is not R but $\omega = 50(1 - \cos\alpha)$. ω is the solid angle subtended by a cone, whose semi-vertical angle is $\alpha = \tan^{-1}(R/u_0)$, expressed as a percentage of 4π steradians. We shall call this curve $D = f(\omega, r, z)$ as the transmission curve for the point (r, z). The transmission curves for two points having the same value of z cannot intersect as that would imply that an electron of a given value of D and angle of emission is passing through two points at the same value of z. The two curves in fig. (7) correspond to the extreme points of the ring. Through a point within these two curves will pass a transmission curve corresponding to a point (r, z_0) where r lies between r_1 and r_2. So such a point will represent an electron which is transmitted by the ring opening. The ordinate of the point gives the value of D of the electron and its abscissa will give the angle of

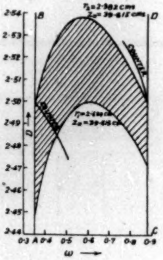


FIG. 7

emission. The points above or below the region between the two curves represent electrons which are stopped by the ring baffle. The action of the baffles limiting the angle of emission is indicated by the lines $\omega = 50(1 - \cos\alpha_1)$ and $\omega = 50(1 - \cos\alpha_2)$ (lines AB and CD). The region within the two curves and the two lines contains representative points of transmitted electrons. If a line $D = D_1$ is drawn, the points on it in the region of transmission give the allowed values of angle of emission. The choice of abscissa has been such that the length of the line intercepted within the region of transmission gives directly the transmission Ω as a percentage of 4π steradians. We get the full line-shape by plotting Ω against D.

The line-shape is shown in fig. 8 (*a*). The peak transmission is 0·52% for $D = 2{\cdot}50$ and the resolution is 1·44%.

Although there is no focusing at the axis, the earlier practice (Deutsch *et al.* 1944) was to place a cylindrical Geiger–Müller counter coaxially in the image space. The counter window served as a discriminating baffle system. The best arrangement is to have the counter window of such dimensions and so placed that it just allows the extreme rays to enter it. The proper place is where the ray corresponding to highest angle of emission is as much below the axis as the other extreme ray is above the axis. This is obtained by solution of the equations

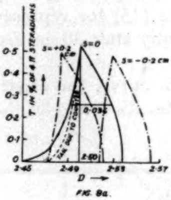


FIG. 8a.

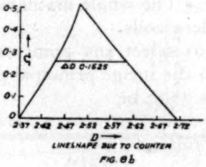


FIG. 8b

$$\frac{r}{R_1} + \frac{z}{v_1} = 1 \quad \text{and} \quad -\frac{r}{R_2} + \frac{z}{v_2} = 1.$$

The solutions z and r give the position and radius of the counter window. This may be called the 'natural counter window'. The line-shape due to a natural counter is shown in fig. 8 (*b*). It is seen that, for the same transmission, ring focusing gives a resolution better than that due to the natural counter by a factor of four.

The result of placing the natural counter behind the ring opening is to cut down the tail on the low D side of the line-shape due to ring alone. The dotted line in fig. 8 (*a*) shows the cut-down low-energy side of the line-shape.

§7. ELIMINATION OF THE TAIL

Perkins and Solbrig (1951) showed that the low-energy tail of the line-shape is greatly reduced if we replace a single outer baffle at z_0 by two outer baffles on two sides of z_0, such that each just transmits the uppermost ray at the respective value of z. We have carried the argument further to show how the tail is completely eliminated by using a continuous system of baffles.

The upper envelope of the rays may be called the caustic curve. Revolution of the curve about the axis of symmetry gives the caustic surface. There will be a family of caustic curves for different values of the parameter D. The coordinate r of a trajectory is given by

$$r=R\left(1-\frac{z}{v}\right)=R\left[1-z\left(\mu+\nu D^2-\frac{1}{u}\right)\right]. \qquad \ldots\ldots(14)$$

The caustic envelope $r_c(z)$ is obtained by elimination of R between the equation

$$(\partial r/\partial R)_{z,D}=0 \qquad \ldots\ldots(15)$$

and the equation (14). The solutions are:

$$\frac{1}{z_c}=(\mu+R\mu')-\frac{1}{u}+(\nu+R\nu')D^2, \qquad \ldots\ldots(15\,a)$$

$$r_c=R^2z_c(\mu'+\nu'D^2). \qquad \ldots\ldots(15\,b)$$

If the angles of emission are limited between α_1 and α_2 the caustic curve is described by (15) between the limits z_2 and z_1 only where z_2 and z_1 are solutions of (15 *a*), obtained on putting $R=R_1$ and R_2 respectively. If $R_2>R_1$, then $z_1>z_2$. From $z=30$ cm to $z=z_2$ the uppermost ray (corresponding to α_2) and beyond z_1 the lowest ray (corresponding to α_1) replace (15) for representation of the caustic curve. The whole discussion is valid only after 30 cm from the central plane of the lens coils.

Now let us select any point on the caustic curve, for electron D_0, whose distance from the image principal plane of the lens is z and let the transmission curve for this point be

$$D=D(R,z,r). \qquad \ldots\ldots(16)$$

We have the relation

$$\left(\frac{\partial D}{\partial R}\right)_{z,r}=-\left(\frac{\partial D}{\partial r}\right)_{z,R}\left(\frac{\partial r}{\partial R}\right)_{z,D}. \qquad \ldots\ldots(17)$$

Since $(\partial D/\partial r)_{z,R}$ is always finite and different from zero, $(\partial r/\partial R)_{z,D}=0$ implies $(\partial D/\partial R)_{z,r}=0$. For the transmission curve in question $(\partial r/\partial R)_{z,D}=0$ when $D=D_0$, so that the transmission curve has a maximum at $D=D_0$ and R given by (14) or (15).

If the upper baffle of the ring is replaced by a continuous system of baffles with circular opening, whose radius at z is given by (15 *a*) and (15 *b*), the single lower transmission curve is replaced by a continuum of curves, everyone of which has a maximum at $D=D_0$. Therefore, the resultant lower transmission curve is just the line $D=D_0$ which is the common tangent to the continuum of curves. The line-shape is shown in fig. 8 (*a*). The tail is replaced by a perpendicular fall. Resolution is improved by a factor of 1·33 and the total spread in D is reduced by a factor of 2·25.

It may be pointed out that the upper point of the natural counter window falls on the caustic curve.

The continuous baffle system has been used by Hey and Latyshev (1946) in the case of the 180° magnetic spectrometer. They obtained line-shape similar to that obtained by us.

For designing the baffle system, a table of r_c against z_c has been prepared with the help of (15 *a*) and (15 *b*). The baffle equation has been plotted in fig. 9.

§ 8. EXTENDED SOURCE

The source is taken in the form of a circular disc with its centre on the axis. Keller *et al.* (1950) have drawn trajectories for meridional as well as skew rays. They have concluded that the $r(z)$ plots of the trajectories of skew rays originating from points on a chord on the source disc are nearly coincident if they are parallel to the meridional plane normal to the chord. In fact, the meridional ray, originating from the point of intersection of a chord and a meridional plane normal to it, may be taken as equivalent to aset of skew rays, parallel to the meridional ray, originating from points along the chord. The consequence of this equivalence is that the total line-shape can be expressed as

$$T(D) = \frac{2}{\pi d^2} \int T(D, s)(d^2 - s^2)^{1/2} ds \qquad \ldots\ldots (18)$$

where $T(D)$ is the transmission for an electron having value D from a circular source of radius d. $T(D, s)$ is the transmission for an electron having value D from a point at a distance s from the centre, by meridional rays alone, and $2(d^2 - s^2)^{1/2}$ is the length of a chord of the source plate at a distance s from the centre. $T(D, s)$ can be calculated with the help of (13) using the relation (4 *a*).

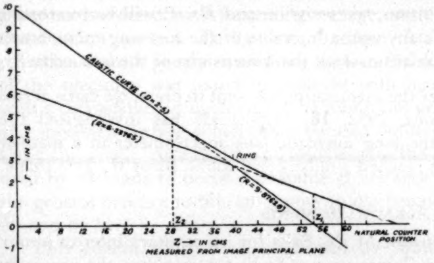


FIG 9

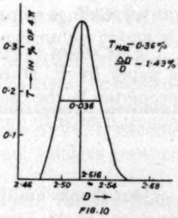


FIG. 10

The calculation of $T(D, s)$ requires parametric relations which are capable of giving information about the variation of u. The relations of other workers are deficient in this respect. In the trajectories calculated by us, u has been varied over a small range only. But this does not weaken our argument. The formulae in the reversible forms are found to be valid over a wide range of values of v. So, by simple interchange of u and v, we can claim that the formulae will be valid for wide ranges of values of u as well.

In fig. 8 (*a*) the line-shapes for $s = 0$ and $\pm 0{\cdot}2$ cm have been plotted for the continuous baffle system. The sharp fall on the low D side of the line for $s = 0$ disappears for non-zero values of s. This departure is very small for small values of s and increases with s. Remembering that the weight factor $(d^2 - s^2)^{1/2}$ decreases as s increases, we may neglect the departure and assume that the sharp edge is retained. The low D side of the line-shape is replaced by a line through the appropriate value of D without altering the half width, in the $s \neq 0$ cases. With this modification, it is seen that the line-shape for non-zero s can be obtained by simply shifting the line for $s = 0$. The fact is expressed as

$$T(D, s) = T(D - \lambda s, 0) \qquad \ldots\ldots (19)$$

with $\lambda = 0{\cdot}10$ cm^{-1}. The total line-shape has been obtained with the help of (18) and (19) and the known line-shape for $s = 0$. A situation like this also occurs for the 180° uniform field spectrometer (Saha 1945).

The optimum condition for source dimension is given by the condition that the line-shape due to $s = -d$ should start where that for $s = +d$ ends. If there is a gap between the two extreme line-shapes, the width of the total line-shape is increased without gain in peak transmission. On the other hand if the two line-shapes overlap, the peak transmission is greatly reduced without comparable improvement of the resolution.

The integrated line shape is presented in fig. 10. The transmission is $0{\cdot}36\%$ at $D = 2{\cdot}516$. The resolution is $1{\cdot}5\%$. The comparatively high resolving power is achieved by the use of the continuous baffle system. For simple ring baffles the resolution is expected to be 2%.

The investigation is not by any means complete. The following problems are yet to be investigated:

(*a*) Dependence of resolution on transmission. The transmission can be altered either by changing the limits on angles of emission or by changing the inner ring baffle. In other words we are required to test whether Frenkel's relation of the long lens spectrometer holds for a short lens.

(*b*) Relation between resolution, transmission and D. It will be interesting to investigate whether there is any optimum value of the focusing momentum.

(*c*) Change of parametric relations with the dimensions of the lens coils.

Note added in proof. After the manuscript was sent to press, we came across a paper by P. Hubert (*Physica*, 1952, **18,** 1129). He has investigated the properties of the caustics for the long magnetic lens spectrometer in a manner almost similar to ours and has reached similar conclusions about the line shape.

ACKNOWLEDGMENTS

Our thanks are due to Professor M. N. Saha for his constant interest in our work and to the Atomic Energy Commission of India for providing the necessary funds.

REFERENCES

BUTT, D. K., 1949, *Proc. Phys. Soc.* B, **62,** 551.

DEUTSCH, M., ELLIOT, L. G., and EVANS, R. D., 1944, *Rev. Sci. Instrum.*, **15,** 178.

HEY, V., and LATYSHEV, G., 1946, *J. Phys. Acad. Sci., U.S.S.R.*, **10,** 446.

KELLER, J. M., KOENIGSBERG, E., and PASKIN, A., 1950, *Rev. Sci. Instrum.*, **21,** 713.

PERKINS, J. F., and SOLBRIG, A. W., Jr., 1951, *Rev. Sci. Instrum.*, **22,** 173.

SAHA, A. K., 1945, *Indian J. Phys.*, **19,** 97.

VERSTER, N. F., 1949, *Appl. Sci. Res.* B, **1,** 363.

ZWORYKIN, V. K., MORTON, G. A., RAMBERG, E. J., HILLIER, J., and VANCE, A. W., 1945, *Electron Optics and the Electron Microscope* (New York: Wiley).

The Diffraction of 150 kv Electrons

By G. I. FINCH*,
H. C. LEWIS† and D. P. D. WEBB
Applied Physical Chemistry Laboratories, Imperial College, London

MS. received 30th March 1953, and in final form 8th July 1953

Abstract. An electron diffraction camera operating at voltages up to 150 kv is described. The advantages of the use of high accelerating potentials are illustrated by results obtained with this camera.

§ 1. Introduction

THE relative advantages in electron diffraction of electron beams accelerated throughout the range between 800 v and 75 kv were examined experimentally in this laboratory between 1933 and 1936. It was found that, for a given transmission specimen, the ratio of elastically to inelastically scattered electrons increased with increasing voltage, with a consequent increase in the transmission pattern clarity and freedom from background. Further, in reflection from a poorly conducting surface the disturbance due to charging-up of the specimen was found to decrease with increasing accelerating potential, with a corresponding decrease in the minimum grazing angle between the beam and specimen surface which could be used without provoking excessive electrostatic charging. As a result, and also because the cold-cathode type of discharge used by us tends to become unstable at voltages much above 65 kv, it became the general practice in this laboratory to use beam-accelerating voltages between 50 and 65 kv.

This general experience, confirmed by Möllenstedt (1946) for the transmission case, suggested that for some purposes it might be useful to use still faster electrons. In the case of electron diffraction by transmission through thin films it seemed probable that a greater range of specimen thicknesses could be tolerated, while in reflection work the depth of penetration below the specimen surface could be increased or diminished beyond that possible at 60 kv, provided that the charging-up difficulties continued to decrease with increasing electron speeds. While Möllenstedt's results suggested, in the case of transmission, a steady improvement with increasing accelerating potential up to and probably beyond 620 kv, practical considerations imposed in our case an upper limit of 200 kv. Even so, relying on Möllenstedt's results, a great improvement could be expected in transmission work. Furthermore, in the case of reflection from poorly conducting surfaces, extrapolation of the results obtained in the range of accelerating potentials already explored by us suggested that on approaching electron velocities corresponding to 200 kv acceleration the depth of penetration of the electron beam below the surface might be considerably reduced, and even confined to the first one or two atom layers.

---

* Now Director of the National Chemical Laboratory, Poona, India.

† Now at St. Benedict's School, Ealing, London.

These views have now been put to the test of experiment, with satisfactory results. It has been found that, in transmission through a specimen of such thickness that the pattern obtained at 60 kv is largely obscured by background scattering, clear patterns are obtained at 150 kv. Also, in reflection from an atomically smooth non-conducting surface such as quartz the least possible angle of incidence can be reduced from about 1° at 60 kv to a few minutes at 150 kv, with the result that the mean depth of penetration of the electron beam below the surface can be reduced from about five or six atom layers at 60 kv to one or two at 150 kv.

§ 2. EXPERIMENTAL

The Electron Diffraction Camera

For purposes of high resolution a two-lens focusing system was adopted. The image of the hot-filament source is brought to a focus a short distance below the upper lens, the reduction in the image size of the emitting source being about tenfold. The second lens, roughly midway between the first lens and the fluorescent screen, focuses the beam on to this screen without further reduction. The electron-optical axes of the lenses are coincident with the electron beam. Both lenses are iron shrouded and are outside the vacuum. Because of the small bore of the tube between the lenses, curvature of the electron beam due to the earth's field is suppressed by mumetal shielding. Two electromagnetic coils with their fields perpendicular to the electron beam

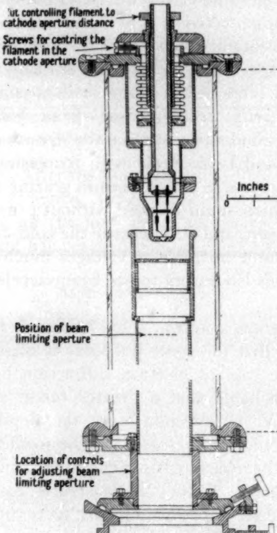

ut controlling filament to cathode aperture distance
Screws for centring the filament in the cathode aperture
Inches
0
1
2
18"
Position of beam limiting aperture
Location of controls for adjusting beam limiting aperture

Fig. 1. The electron gun.

and to each other are situated between the specimen and the second lens; they serve to deflect the beam for exploring the specimen without disturbing the coaxial path of the electron beam through the lower lens. The lower part of the camera, including the specimen holder, is of the standard 'Finch' type (Finch and Wilman 1937 a).

The Electron Gun (fig. 1)

Owing to the instability of the cold-cathode type of discharge above 70 kv, a three-electrode hot-filament source was used. The relative positions of the filament and Wehnelt cylinder, and the bias on the latter are adjustable. The gun was designed to operate at up to 200 kv. The Pyrex glass insulator, however, does not withstand more than about 160 kv continuous working, but could be replaced by a more robust porcelain cylinder.

The High Tension Source

The high-tension source consists of a 110 kv mains frequency transformer working into a Cockcroft voltage doubler circuit designed and constructed by Messrs. Ferranti Ltd. The output of the doubler is fed to the diffraction camera cathode through a 200 kv rectifying valve with an under-run filament which provides against overloading.

§ 3. EXPERIMENTAL RESULTS

3.1. *Transmission*

A mica single crystal, about 5000 Å thick, was examined at 60, 100 and 140 kv with the beam roughly parallel to the *c* axis direction of the crystal. A pinhole diaphragm placed close above the specimen restricted the electron beam to the same uniformly thick part of the crystal flake. The resulting patterns are shown in figs. 2, 3 and 4 (Plate I), camera length $L=48$ cm. The gain in clarity and detail and hence significance of the pattern with increasing electron speed is striking.

3.2. *Reflection from a Single Crystal Surface*

(*a*) A rock salt single-crystal cleavage surface gave, with a 60 kv beam, the pattern shown in fig. 5 (Plate I), $L=48$ cm. Although a decharger was used, the blurring effects of charging-up in the beam are prominent even though the angle of incidence was as much as about 1°. With a 120 kv beam, however, at the same angle of incidence, the pattern was clear in all its details, fig. 6 (Plate I), $L=48$ cm.

(*b*) Owing to the existence of steps, a rock salt cleavage face is not suitable for examining the effective depth of penetration of the electron beam below the surface at low grazing angles. For this purpose we have found that mechanically polished single-crystal surfaces of quartz, corundum, spinel and rutile are satisfactory. With care, atomically smooth surfaces of sufficient extent can be developed by polishing with alumina and water on Selvyt cloth. Results obtained at 60 and 140 kv with a polished quartz crystal are shown in figs. 7, 8 and 9 (Plate II), $L=48$ cm. At 60 kv and with a decharger the smallest useful angle of incidence was about 1°. A veiling of the pattern near the shadow edge was evident and is suggestive of a non-crystalline surface film (fig. 7). Retaining the same angle of incidence but increasing the beam voltage to 140 kv greatly reduced this veiling (fig. 8). At this voltage, because of the reduced effect of charging-up, the angle of incidence could be reduced to about 10′. When this was done, the veiling returned, and then clearly consisted of very diffuse haloes (fig. 9).

It is extremely unlikely that these haloes are due to contamination. The only likely source of extraneous matter is that due to deposits from hydrocarbon or other vapours on the specimen during examination in the electron beam. This can be very troublesome in electron microscopy where the electron beam is concentrated on the specimen, but in electron diffraction the energy flux at the specimen is very much less and such deposits are not normally observed. A fresh cleavage of muscovite mica, which has an almost perfect atomically smooth surface, was examined at very low grazing angles of incidence and no haloes were observed. Furthermore, no difference was observed between patterns taken as soon as possible after evacuating the camera and patterns taken after the specimen had been irradiated by the decharger and electron beam for three minutes. In addition, the halo pattern obtained from the polished quartz agrees with that to be expected theoretically from amorphous SiO_2. This confirms that the polish layer is amorphous SiO_2, although extremely thin.

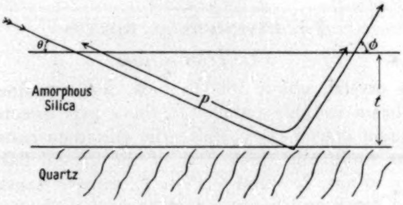


Fig. 10.

Because the pattern due to the underlying quartz single crystal is visible with the halo pattern, it is possible to set an upper limit to the thickness of this layer. If t (fig. 10) is the thickness of this layer, θ and ϕ the angles of incidence and emergence respectively of the electron beam, and p is the greatest path length of electrons which contribute to the quartz single-crystal pattern, then $t = p \sin\theta \sin\phi / (\sin\theta + \sin\phi)$. According to Thomson (1934), the mean free path for ionization in a solid is 400 Å at 30 kv and is roughly proportional to the voltage. There will also be inelastic collisions which only result in excitation of the atoms without ionization, and the effect of these will be to reduce the effective mean free path. Using $\theta = 10'$ and $\phi = 1^\circ\ 50'$ (obtained from fig. 9), we obtain the value of five ångströms for the thickness of the amorphous layer. Results obtained by Finch and Wilman (1937 b) and Lucas (1951) show that satisfactory estimates can be made of the thickness of such thin layers.

Results similar to those obtained with polished quartz have been obtained with a polished corundum crystal, the amorphous polish layer proving to be again only about one or two atom layers thick. In the case of the rutile single crystal examined the amorphous polish layer was, however, much thicker. Thus while an unpolished rutile single-crystal surface gave the pattern shown in fig. 11 (Plate II), it yielded after polishing only a halo pattern, no trace of Kikuchi lines being visible even when the angle of incidence was increased to beyond 4°. Evidently the amorphous polish layer formed on rutile is comparatively thick and has a minimum thickness of about 70 Å.

This high-voltage electron diffraction camera has now been in constant use since early 1951. Our general experience may be summed up by saying that, in transmission, a far wider range of specimen thicknesses can be tolerated than in the case of the more normal 60 kv technique. Furthermore, many pseudo-reflection specimens which yield no recognizable pattern at the lower voltage (due to their surface asperities being too thick for the electron beam to pass through without much loss of energy), give clear and significant patterns at 140–150 kv. Finally, in reflection from atomically smooth non-conducting surfaces, the depth of beam penetration below the surface can be reduced to little more than the top atom layer, or increased far beyond that achievable with 60 kv electrons.

ACKNOWLEDGMENTS

One of us (G.I.F.) wishes to thank the Central Research Fund of the University of London for providing the Ferranti high tension equipment; the other co-authors (H.C.L., D.P.D.W.) wish to thank the Department of Scientific and Industrial Research for grants.

REFERENCES

FINCH, G. I., and WILMAN, H., 1937 a, *Ergebn. exakt. Naturw.*, **16,** 353; 1937 b, *Trans. Faraday Soc.*, **33,** 337.
LUCAS, L. N. D., 1951, *Proc. Phys. Soc.* A, **64,** 943.
MÖLLENSTEDT, G., 1946, *Nachr. Wiss. Göttingen*, **1,** 83.
THOMSON, G. P., 1934, *Phil. Mag.*, **18,** 640.

An Optical Method of Observing Stress Relaxation in Transparent Solids

By S. M. CRAWFORD

Imperial Chemical Industries Limited, Butterwick Research Laboratories,
Welwyn, Herts.

MS. received 11*th June* 1953

Abstract. The relaxation of birefringence produced in two organic glasses after they have been sheared in a modified coaxial cylinder viscometer has been followed at different temperatures by the use of a photocell. For the low rates of loading used it has been found that the relaxation of both the glasses can be represented by a single exponential function, which shows that there is one predominant relaxation mechanism. The relaxation times were measured at different temperatures and cover a range from about 20 seconds to ½ millisecond. The viscosity has also been measured over the same temperature range and, from the values of the viscosity and relaxation time at a given temperature, the rigidity modulus of elasticity has been determined. Of the two materials used, one, glycerol sextol phthalate, was more chemically associated than the other, 2′-hydroxy-2:4:4:6:5′-pentamethylflavan. This difference was reflected in the temperature variation of the quantities measured.

---

§ 1. Introduction

When many materials are stressed and held at constant strain, the stress is observed to relax with time. In the simplest case, this relaxation is represented by an equation of the form $\sigma = \sigma_0 e^{-t/\tau}$, where σ_0 is the value of the stress at zero time and σ its value after time t. The constant τ is known as the relaxation time and gives the time for the stress to relax to 1/e of its initial value.

Investigation of the visco-elastic properties of high polymers has shown that their behaviour cannot be explained in terms of a single relaxation time, but any relaxation phenomenon is the sum of a large number of superimposed relaxation mechanisms (Alfrey 1948). Attempts have been made to explain the different times of relaxation in terms of the molecular structure of these materials, but the problem is complicated by the fact that the constitution of such polymers and the nature of the intermolecular forces is not rigidly defined.

There are some amorphous materials, however, which have a well-defined molecular constitution with no entanglement of molecules, such as occurs in high polymers. These may be expected to have a simpler relaxation time spectrum and therefore, by working first with such materials and then with materials intermediate between these and high polymers, it might be hoped to gain some insight into relaxation mechanisms.

The purpose of this work, therefore, was to follow directly the stress relaxation of materials with comparatively simple molecules after they had been statically loaded at different temperatures. A single exponential function would fit the relaxation curve if the material had only one relaxation time, while it would allow of analysis into several exponential functions if there were several relaxation mechanisms.

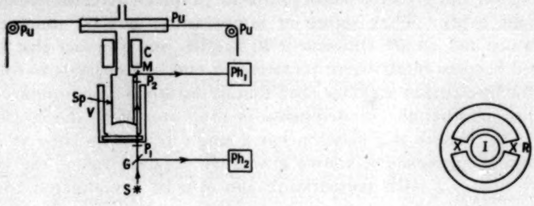
Amorphous organic materials of known composition were used whose viscosity increased continuously from the liquid to the solid state. For convenience, these will be referred to as organic glasses and in these experiments the glasses used were 2′-hydroxy-2:4:4:6:5′-pentamethylflavan (Baker, Curtis and McOmie 1952) and glycerol sextol phthalate (a mixed glycerol-cyclohexanol ester of phthalic acid). They soften at temperatures not far removed from room temperature and so are convenient to handle, and the fact that they are transparent and become birefringent on stressing makes it possible to follow the relaxation of the birefringence rather than that of the stress; preliminary experiments having shown that the birefringence is proportional to the stress in the region employed. Where the material has a single relaxation time at any one temperature and the viscosity is known over a temperature range, the variation of the rigidity modulus with temperature can also be investigated, since the relaxation time τ is given by $\tau = \eta/n$ where η is the viscosity and n is the rigidity modulus of elasticity.

§ 2. EXPERIMENTAL

The original intention was to shear the material in a coaxial cylinder viscometer by rotating the inner cylinder, and to bring this cylinder to rest by letting it strike a rigid stop. The stress in the glass would thus relax and be taken up by the stop. It was found, however, that the deformation of the stop under the applied load was of the same order as the shear produced in the material. The relaxation time observed, therefore, was the deformation time of the steel under the action of an increasing load, whose rate of increase to the steady value of the applied load was governed by the viscosity of the material.

In order to obviate this difficulty, it was decided to use an arrangement in which the effect of the deformation of the steel could be eliminated. The method involves shearing the material and a steel member in parallel with it. Initially, both the material and the steel are under stress, but, as the stress relaxes in the material, it is taken up by the steel. The relaxation process observed is that of the steel–glass combination, but by repeating the measurements with different amounts of material the effect of the steel can be allowed for and the relaxation time of the material determined. This method has the additional advantage that any flow birefringence which occurs will be negligible. The birefringence will, thus, be entirely due to the stress in the material.

The apparatus is shown diagrammatically in fig. 1. The material was sheared between two coaxial cylinders with a thin steel bar connecting the inner and outer cylinders. The torque was applied to the inner cylinder by pulleys and weights and end effects were reduced by trapping an air bubble beneath the inner cylinder. The whole apparatus was placed in an enclosure the temperature of which could be varied and controlled by a thermostat. Cooling was obtained by means of a refrigerating unit. The temperature was measured with a thermocouple whose hot junction was embedded in the outer cylinder. This cylinder had a glass bottom so that a beam of light could be passed up through the annular space between the cylinders. The beam was plane polarized before entering the material and allowed to fall on a second polaroid, crossed with the first, after leaving the material. A photocell was used to detect the light transmitted by the analyser when the material was birefringent, and, in order that it should respond to small differences in refractive index, a ciné projector lamp was used as a high intensity source of light. The fraction of the incident light transmitted in these experiments was very small so that the phase differences produced were such that the output from the photocell could be taken as linearly proportional to the birefringence.



(a) (b)

Fig. 1 (a). Experimental arrangement. V, viscometer; Sp, specimen of material; S, light source; G, glass plate to split light beam; M, plane mirror; P_1, P_2, crossed polaroids; Ph_1, Ph_2, photocells; Pu, pulleys; C, bearing for inner cylinder.
(b). Plan of top of container: I, inner cylinder; R, rim of containing vessel; XX, double steel sector in parallel with specimen of material.

Although the projection lamp was run from a constant voltage mains transformer, it was found that the variations due to fluctuations in light intensity still appeared in the output from the photocell. It was arranged, therefore, to expose a second photocell to the light source and balance its output against that of the first. The circuit is shown in fig. 2 and is a modification of that described by Brown and Hickson (1950). The output from the photocell was fed to an electrometer triode for amplification and the galvanometer, placed across the balanced bridge, recorded any change in intensity of light due to birefringence effects.

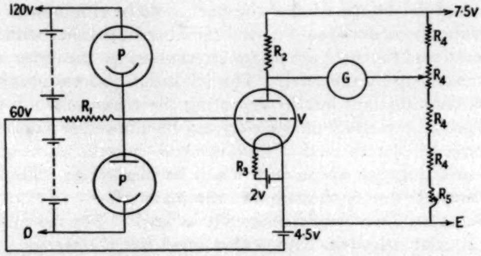


Fig. 2. Photocell circuit. P, photocell, VA 26T; V, electrometer triode, ET_1; $R_1 = 10\,000$ MΩ; $R_2 = 10\,000$ Ω; $R_3 = 15$ Ω; $R_4 = 1000$ Ω; $R_5 = 200$ Ω.

A vibration galvanometer with a period of 33 msec was used so that short relaxation times could be measured. The upper time limit of the method was fixed by the stability of the apparatus and, in practice, relaxation times of the order of 1 minute were the longest to be measured. The galvanometer spot was focused on to a variable speed recording camera with a continuous paper feed. The time scale was obtained by focusing a second spot on to the camera and interrupting the beam at regular intervals. The synchronous motor of an electric clock was used for this purpose.

The bridge was first balanced with no light, and then, with the projector lamp on, the polaroids controlling the light to the compensating photocell were adjusted to give zero deflection of the galvanometer. The camera was switched on and a torque suddenly applied to the specimen. Since the relaxation time was found to be the same whether the system was loaded or unloaded, it was convenient to unload the specimen by fusing a wire supporting the weights. In this way rapid application of the stress was obtained without vibration.

For relaxation times of the same order as, and less than, the natural period of the galvanometer, a slightly different technique was adopted. A single photomultiplier was used to detect the birefringence and its output was fed to the d.c. amplifier of a double-beam cathode-ray tube. If the appropriate time base was used, the relaxation phenomenon could be displayed on the tube and photographed. A timing trace was also obtained either from the a.c. mains or, at faster sweep times, by a crystal oscillator. A relay was arranged to unload the system as the camera button was pushed and, as the wire supporting the weights become hot and the resistance in the loading circuit increased, a second relay triggered the beam on the tube. The lower time limit of the method is set by the inertia of the system, and relaxation times much shorter than 1 msec could not be measured, since with this apparatus, 1 msec appears to be the order of time required for the birefringence to reach its maximum value when the system is unloaded.

Experiments were carried out at different temperatures with a known length of material in the annular gap and were then repeated over the same temperature range with a different amount of material.

In order to measure the viscosity, a simple coaxial cylinder viscometer was used where the inner cylinder rotated under the action of a couple. The gap width was rather larger than would be normally employed, in order to transmit the light for the birefringence experiments, and was about 5 mm, the outer cylinder having a radius of 2·5 cm. The length of the inner cylinder was approximately 8 cm. A couple was applied to the inner cylinder by means of a pulley of radius 6·5 cm, and the velocity of rotation was observed for different loads at each temperature. The viscosity was then determined from these results and the dimensions of the apparatus.

§ 3. THEORY OF THE METHOD

The system under stress can be conveniently considered in terms of a mechanical model consisting of a spring in parallel with a Maxwell element, which itself consists of a spring and dashpot connected in series (fig. 3). The Maxwell element represents the glass and the spring represents the steel rod connecting the inner and outer cylinders.

When a stress is first applied to this system, it is divided between the two arms in the ratio of the stiffness of the two springs S_1 and S, the stiffness being the ratio between the stress and the extension it produces. As the viscous element, with constant η_1^*, equal to the ratio between the stress and the velocity it produces, comes into play, the stress in this arm relaxes until eventually all the stress is taken up by the spring of stiffness S. This apparent time of relaxation τ_1 depends on the values of S, S_1 and $\eta_1{}^*$ and is given by:

$$\tau_1 = \eta_1{}^* \left(\frac{1}{S} + \frac{1}{S_1} \right). \qquad \text{......(1)}$$

The values of $\eta_1{}^*$ and S_1 depend on the geometry of the specimen as well as its physical properties. Thus, if a couple G_1 is applied to the inside of a hollow cylinder of material of rigidity modulus n, so that it is twisted equally throughout its length l_1, then the angle of twist θ_1 is given by

$$G_1 = 4\pi n \frac{a^2 b^2}{b^2 - a^2} l_1 \theta_1, \qquad \text{......(2)}$$

where a and b are the internal and external radii respectively. Therefore

$$S_1 = G_1/\theta_1 = 4\pi n \frac{a^2 b^2}{b^2 - a^2} l_1. \qquad \text{......(3)}$$

Similarly

$$\eta_1{}^* = G_1/\Omega_1 = 4\pi\eta \frac{a^2 b^2}{b^2 - a^2} l_1, \qquad \text{......(4)}$$

where η is the viscosity of the material and Ω_1 the angular velocity produced by the couple G_1. Therefore, in (1),

$$\tau_1 = 4\pi\eta \frac{a^2 b^2}{b^2 - a^2} l_1 \frac{1}{S} + \frac{\eta}{n}. \qquad \text{......(5)}$$

But the relaxation time of the material is given by

$$\tau = \eta/n, \qquad \text{......(6)}$$

so that

$$\tau_1 = \tau + 4\pi\eta \frac{a^2 b^2}{b^2 - a^2} l \frac{1}{S}. \qquad \text{......(7)}$$

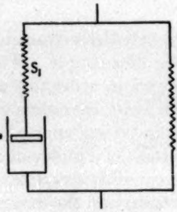


Fig. 3. Mechanical model of system.

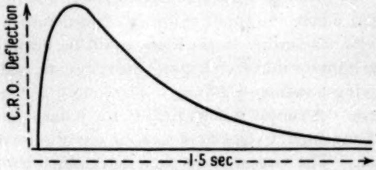


Fig. 4. Trace obtained with 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan at 14·3°c.

If the stiffness of the steel is very large compared with that of the organic glass, then the last term in eqn. (7) can be neglected, but in practice this could not be realized because the amount of birefringence resulting from the small stress in the glass was found to be too small for the relaxation to be followed.

In order to eliminate S, therefore, the experiment is repeated with a different amount of material at the same temperature. The relaxation time τ_2 is then given by

$$\tau_2 = \tau + 4\pi\eta \frac{a^2 b^2}{b^2 - a^2} l_2 \frac{1}{S}. \qquad \text{......(8)}$$

Thus, from eqns. (7) and (8),

$$\tau=\frac{l_2\tau_1-l_1\tau_2}{l_2-l_1}=\frac{\tau_1-(l_1/l_2)\tau_2}{1-l_1/l_2}. \qquad \ldots\ldots(9)$$

The final result is thus independent of the dimensions of the apparatus and depends only on the ratio of the lengths of material employed and, in practice, this ratio was approximately $l_1/l_2=\frac{1}{2}$.

When the stress is removed, it can be shown that, for such a system, similar effects will occur and the birefringence will show the same relaxation time as when the stress is applied.

§ 4. Results

Figure 4 shows a typical trace obtained with hydroxy-pentamethylflavan and fig. 5 gives the corresponding graph of the logarithms of the deflections against time. Similar graphs have been obtained for both substances throughout the temperature ranges covered and the fact that these plots are linear shows that for these low rates of loading the materials have a single relaxation time or, at least, one relaxation mechanism which is predominant.

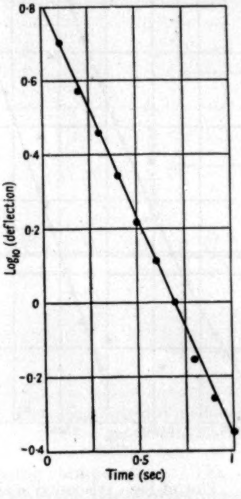


Fig. 5. Plot of $\log_{10}$ (deflection) against time, from trace in fig. 4.

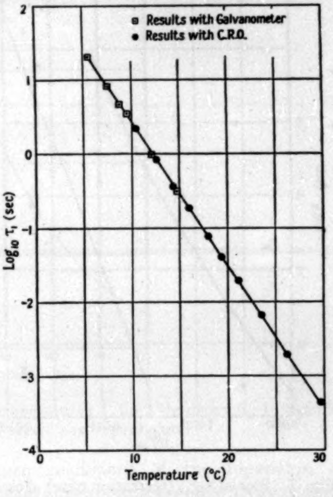


Fig. 6. Plot of $\log_{10}\tau_1$ against temperature for 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan.

The relaxation times of the steel–glass combination, τ_1 or τ_2, were determined at each temperature from the graphs obtained, as in fig. 5. Figure 6 shows the graph of $\log\tau_1$ against temperature for hydroxy-pentamethylflavan and it is seen that the results for the two slightly different techniques agree with one another. A similar graph was obtained for τ_2 and, from these, the variation with temperature of the relaxation time of the material τ was determined by means of eqn. (9).

When log τ was plotted against temperature, the plot was found to be linear for the flavan compound. Figure 7 shows the graphs of log τ against the reciprocal of the absolute temperature for the two materials, and it is seen that the plot is still approximately linear for the flavan compound over the narrow temperature range covered. The relaxation time therefore obeys an equation of the form, $\tau = \tau_0 e^{A/T}$, where τ_0 is the value of τ when T is infinite and A is a constant. Any process which obeys such an equation has an energy of activation in appropriate units of $A/\boldsymbol{R}$ per mole, where $\boldsymbol{R}$ is the molar gas constant. The gradient of the log τ against $1/T$ graph gives A, from which, in the case of hydroxy-pentamethylflavan, an activation energy of 76 kcal per mole is obtained. The graph for glycerol sextol phthalate is not linear so that the activation energy varies with temperature. It is of the order of 66 kcal per mole at the lowest temperatures and 44 kcal at the higher temperatures.

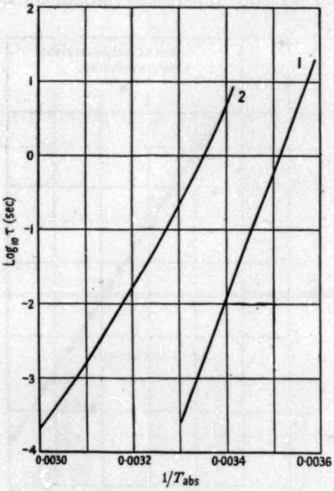


Fig. 7. Plot of $\log_{10}$ (relaxation time) against the reciprocal of absolute temperature: 1, for 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan; 2, for glycerol sextol phthalate.

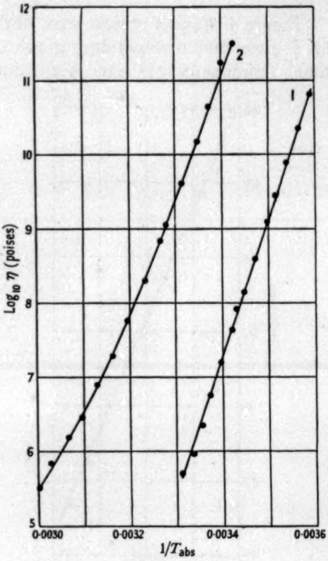


Fig. 8. Plot of $\log_{10}$ (viscosity) against the reciprocal of absolute temperature: 1, for 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan; 2, for glycerol sextol phthalate.

The value of the viscosity was found to be independent of the load used and the curves showing log (viscosity) against reciprocal of the absolute temperature are shown for the two materials in fig. 8. Here again, that for the flavan compound is linear and gives an activation energy of some 86 kcal per mole. The curve for glycerol sextol phthalate gives activation energies ranging from 76 to 56 kcal per mole.

Since the materials have a single relaxation time, a rigidity modulus *n* can be determined at each temperature from the values of τ and η determined as described, since $\tau = \eta/n$, but the values are only significant so long as η is independent of frequency. Measurements of the apparent viscosity over a wide range of frequencies (J. J. Benbow 1953, to be published) have shown that the viscosity is almost independent of frequency for hydroxy-pentamethylflavan over the range of relaxation times covered, but it varies considerably in the case of glycerol sextol phthalate. Figure 9 shows the graphs, plotted against temperature, of the logarithms of the rigidity modulus, calculated using the viscosity measurements obtained from the results shown in fig. 8, and it is seen that the plot is linear for hydroxy-pentamethylflavan and shows decreasing elasticity with increasing temperature. The temperature coefficient of the logarithm of elasticity in the case of glycerol sextol phthalate appears to decrease with increasing temperature.

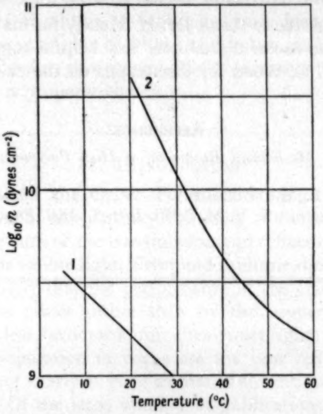


Fig. 9. Plot of $\log_{10}$ (rigidity modulus) against temperature: 1, for 2′-hydroxy-2 : 4 : 4 : 6 : 5′-pentamethylflavan; 2, for glycerol sextol phthalate.

§ 5. Discussion

The most important result which emerges from these experiments is that, for these materials, there appears to be only one mechanism responsible for the relaxation phenomena observed at low rates of loading. Any others which occur, must occur very rapidly in comparison with those observed. Measurements of the dynamic shear modulus and viscosity of these materials carried out over a wide range of frequencies and at different temperatures have confirmed this result (J. J. Benbow 1953, to be published).

The chemical nature of glycerol sextol phthalate is such that it might be expected to be associated since there are several free hydroxyl groups, and the fact that the rates of fall with temperature of the logarithms of the viscosity and relaxation time become less at higher temperatures tends to confirm this. The comparatively rapid reduction in the logarithm of elasticity with temperature at first, followed by a smaller rate of reduction, may also be attributed to decreasing association. Hydroxy-pentamethylflavan has only one hydroxyl group and it is in such a position that association is less likely to occur than in glycerol sextol phthalate (Baker, Curtis and McOmie 1952). The high activation energy compared with that of glycerol sextol phthalate may possibly be attributed to the greater rigidity of the molecule.

It would seem, then, that for these substances the relaxation processes observed under these conditions are the relaxations of the stresses between the molecules or groups of molecules. It will be of interest to carry the work further using materials of a more complex structure.

ACKNOWLEDGMENTS

The author would like to thank Dr. H. Kolsky for his guidance in the work, Mr. J. J. Benbow for useful discussions and helpful suggestions of a practical nature and Dr. D. J. C. Wood for discussions on the chemical nature of these materials.

REFERENCES

ALFREY, T., Jr., 1948, *Mechanical Behaviour of High Polymers*, Section B (New York : Interscience).

BAKER, W., CURTIS, R. F., and MCOMIE, J. F. W., 1952, *J. Chem. Soc.*, 1774.

BROWN, A. F. C., and HICKSON, V. M., 1950, *Brit. J. Appl. Phys.*, **1**, 39.

Reflectivity of Thin Aluminium Films and their Use in Interferometry

BY J. C. BURRIDGE, H. KUHN AND ANNE PERY
The Clarendon Laboratory, Oxford

MS. received 24*th June* 1953

Abstract. The coefficients of transmission and reflection of thin aluminium films deposited on silica by vacuum evaporation were measured for wavelengths between 4200 and 2600 Å. The results were applied to the calculation of the width and intensity of Fabry–Perot fringes obtainable with aluminium coated plates. Measurements of widths of heterochromatic fringes gave upper limits of the reflectivity at wavelengths of 2400 Å and below, and showed that 1/18 of one order can be resolved in this range without excessive loss of intensity if the thickness of the films is properly chosen.

§1. INTRODUCTION

THE performance of the Fabry–Perot etalon depends critically on the optical properties of the partially reflecting coatings; it is the better the more closely the sum of the transmission and reflection coefficients, $T+R$, approaches unity. For visible light, silver and multiple dielectric films approach this ideal value so closely that the performance of the etalon is usually limited by the flatness of the plates rather than by the properties of the coatings. Conditions are much less favourable for ultra-violet light, for which aluminium films deposited by evaporation *in vacuo* are the best reflectors available; but their performance is not nearly as good as that of silver films in the visible and is far less well known. Of the large number of publications on the properties of aluminium films, only very few are concerned with the optical properties of the type of film which is of interest for interferometric spectroscopy, namely films transmitting about 2 to 20% of the light.

The present work has been undertaken with the aim of providing some of this missing information. It forms an extension to the ultra-violet of the work by Kuhn and Wilson (1950) on optical properties of silver films.

§2. THE PRODUCTION OF THE FILMS

Previous experience had shown that the vacuum conditions during evaporation affect the optical properties of aluminium films much more critically than those of silver films. A more powerful mercury diffusion pump was therefore fitted to the evaporation chamber described before (Kuhn and Wilson 1950); its estimated speed, with liquid air trap, was 100 l. sec^{-1}. The aluminium, which was evaporated from a tungsten filament, was usually of 99·99% purity. In most other respects the procedure of cleaning of the targets, of discharge treatment and of evaporation was the same as for silver. Three or four optically flat silica plates, at distances of 22 to 28 cm from the filament, were coated simultaneously in each evaporation. The pressure was measured by means of a Penning gauge whose calibration was approximately known.

Before evaporation the filament was heated in order to outgas the system. When the aluminium bead started evaporating, the pressure rose to about 5×10^{-5} mm Hg. The shutter between bead and target was opened a little later when the pressure had decreased again, so that the films were generally deposited at pressures between 1 and 3×10^{-5} mm, depending largely on the previous outgassing. The release of gas during the evaporation is typical of aluminium and was not observed for silver. A beam of light passing through the tank and focused on a selenium cell allowed the density of the film to be measured during evaporation so that films of any specified transmission could be produced. This measurement was not as accurate as the measurements described below, which were carried out after the target had been removed from the tank.

§ 3. THE OPTICAL MEASUREMENTS

For wavelengths from 4200 to 2600 Å the values of T and R were measured directly by means of the simple apparatus described by Kuhn and Wilson. An ultra-violet sensitive photomultiplier type RCA I.P.28 was used and the incandescent source was replaced by a hydrogen discharge tube. Four bands of wavelengths of widths about 300 Å were isolated by means of the combinations of filters shown in the table. In view of the very gradual change of T and R with wavelength these bands were found to be narrow enough for the present purpose:

4200 Å: Wratten filters 36 and 43.
3600 Å: Wood's glass.
3200 Å: dense film of silver on silica.
2600 Å: chlorine, bromine and Chance's glass OX7.

The errors of measurement, which were mainly due to the instability of the hydrogen discharge tube, were about 0·01 in R and 0·005 in T. The linearity of the galvanometer deflection was checked by means of the inverse square law, with the use of a point source of visible light.

§ 4. RESULTS OF THE DIRECT MEASUREMENTS

One of the most striking properties of aluminium films is the marked increase of light transmission with decrease of wavelength. This fact, which has of course been qualitatively found by other observers, is shown in fig. 1, where is plotted the ratio of T for a given wavelength to its value for 3600 Å. Each point is the average of measurements of 20 films whose values of T at 3600 Å varied from 0·01 to 0·20. No significant variation T/T_{3600} with T was found in this range; the curve can therefore be used in most practical cases for deriving values of T of an aluminium film from the value of T measured at one wavelength.

The main results of the direct measurements are contained in figs. 2 to 5, showing $R+T$ as function of T. Each point refers to a different film and is the average of several measurements. The scatter of the points about a smooth curve is distinctly greater than the errors of measurement and shows a genuine scatter in the properties of the films. Films deposited simultaneously generally differed less from one another than from other films. This indicates that conditions of evaporation and discharge treatment are more important than conditions of cleaning. The points plotted are, of course, complete sets of experiments and not films selected according to their properties.

The measurements in figs. 2 to 5 were taken within a few days of the evaporation. For a number of films the change of the optical properties with time was studied by repeated measurements from a few hours to three months after evaporation while the films were stored in moderately clean air, protected from dust. The changes were always small, certainly after the first day or two, and especially small for dense films. There was a tendency for T to increase by amounts under 1%, while R decreased by slightly more. These results are perfectly compatible with those of Cabrera (1944), who observed considerable increases of T and decreases of R only for thinner films and only during the first few hours after evaporation. The changes are probably due to the formation of a transparent oxide film (Sennett and Scott 1950).

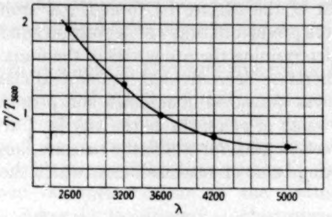
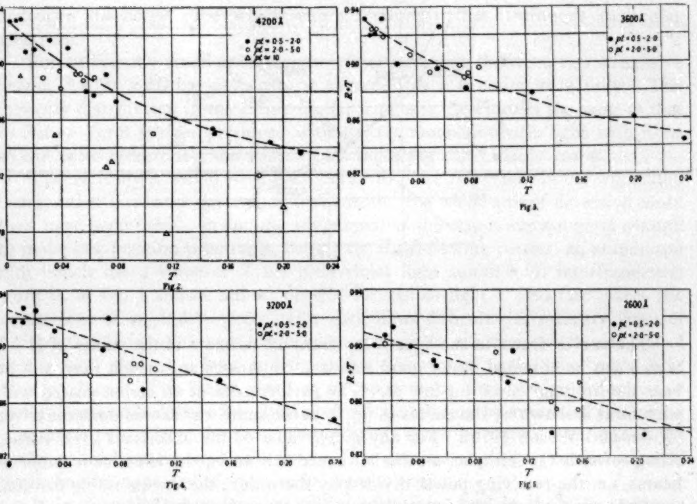


Fig. 1.



Figs. 2–5. p in mm Hg $\times 10^{-5}$; t in minutes.

Various observers have reported that slow deposition of aluminium films produces low reflectivity. Sennett and Scott attribute this fact to differences in the size of the crystals. Though this difference no doubt exists, it appears more likely to us that the variation of the reflectivity with speed of evaporation is due largely, or even entirely, to the influence of the residual gases during the deposition of the film. While a metal film is being formed the percentage of foreign gas atoms occluded in the metal can be expected to depend on the relative rates of bombardment of the surface by foreign gas atoms and metal atoms. If this view is correct the product of time t of deposition and foreign gas pressure p should be the factor determining the properties of the films. The points plotted in figs. 2 to 5 have therefore been divided into several classes according to the value of pt, where t was measured in minutes and p in units of 10^{-5} mm Hg. The influence of the value of pt on the reflectivity is clearly marked at 4200 Å, less so for the other wavelengths, for which the data are, however, less complete. It appears that a limiting value of pt exists below which the reflectivity does not increase.

Whenever the evaporation was interrupted for a few minutes without any admission of air, the resulting value of $R+T$ was well below its normal value.

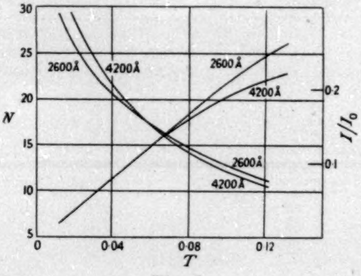


Fig. 6.

The curves drawn in the figures represent an estimate of the values of $R+T$ which can be obtained under good standard conditions and which may well be near the limiting values for low pt. The performance of an etalon coated with such films is shown by the curves in fig. 6, in the same way as was done for silver (Kuhn and Wilson 1950). For any given value of the abscissa T, the curves rising towards the left give, on the left coordinate scale, the effective number of beams, i.e. the resolving power divided by the order; the curves rising towards the right give, on the right coordinate scale, the value of the intensity relative to the intensity in the absence of an etalon, in the maxima of the fringes obtained with perfectly monochromatic light. While for denser films ($T<0{\cdot}06$) the performance for ultra-violet light is slightly worse than that for visible light, thinner films show the reverse behaviour. This is presumably due to interference effects within the film.

In comparing these results with those reported by Crawford, Gray, Schawlow and Kelly (1949) and by Sennett and Scott (1950) it has to be considered that these authors investigated only a few films and probably did not use very low values of the factor *pt*. Our values of $R+T$ are noticeably higher than those reported by these and earlier authors.

Sennett and Scott found that, for silver also, slow evaporation produces low values of reflectivity and, as in aluminium, they ascribed this effect to differences in crystal size. It seems likely to us that in silver too the influence of the residual gas is partly or wholly responsible for the time effect, but the limiting value of *pt* for silver appears to be so high that, in ordinary evaporation practice, it is easy to get the highest possible reflectivity at a pressure of about 5×10^{-5} mm Hg unless the time of evaporation of the silver is unduly prolonged.

§ 5. MEASUREMENTS AT SHORTER WAVELENGTHS BY AN INDIRECT METHOD

The good performance of aluminium films for wavelengths of 2600 Å made it highly desirable to extend the measurements to shorter wavelengths. The decreasing sensitivity of the photomultiplier and the lack of suitable filters would have made this difficult by means of the direct method. An indirect method was therefore used, based on measurement of the width of heterochromatic interference fringes (Edser–Butler fringes). Aluminium-coated etalon plates were set up with three loops of $10\,\mu$ copper wire as spacers. The etalon was placed in front of the slit of a quartz spectrograph illuminated by means of a hydrogen discharge tube which emitted strongly the continuous molecular spectrum.

For sufficiently small aperture, the phase difference between two successive beams, $\delta=(2\pi\times2D\cos\phi)/\lambda$, can be regarded as a function of λ only. The intensity distribution in the Edser–Butler fringes is then ideally the same as that in Fabry–Perot fringes produced with strictly monochromatic light with films of the same reflectivity, the variable $1/\lambda$ replacing the variable $\cos\phi$.

Theoretically it would have been better to place the etalon in the collimated beam, either inside the spectrograph or outside; the whole area of the etalon could then have been used. In fact the adjustment of the etalon was not good enough to make this position preferable, largely for the following reason: an aluminium film which has a value of T for ultra-violet light suitable for interferometric work is, as fig. 1 shows, almost opaque for visible light. This fact makes the adjustment of an etalon coated with such films difficult. A mercury lamp of 125 watts was focused on a 1 mm wide aperture in a screen which was placed in the focal plane of a lens, and the etalon was viewed against the bright light thus produced. The etalon acted as an interference filter of high order and in general appeared dark. By adjustment of the spacing by means of the spring pressure on the plates the transmission bands could be shifted until one of the mercury lines was transmitted strongly. The pressure studs were further adjusted until the film appeared of uniform brightness or tint.

It is clear that imperfections in the adjustment and in the silica surfaces, imperfect collimation and finite slit width will all increase the width of the fringes. The value of R calculated from the measured fringes will therefore give a lower limit for the actual reflectivity. Figure 7 shows a photometer tracing of some of the fringes thus obtained. The distance between orders is about 30 Å in this range. The horizontal arrows indicate the level of half-intensity, as derived from intensity marks. Measurements of this tracing showed that the half-value width is about 1/18 of one order, corresponding to a lower limit for R of 0·84 for 2400 Å. At 2300 Å a width of 1/15 of one order was found, giving $R>0{\cdot}81$.

An extension to shorter wavelengths would have required the use of special photographic plates. This extension, and also a more complete, quantitative investigation giving values of T and more accurate values of R, could not, unfortunately, be undertaken.

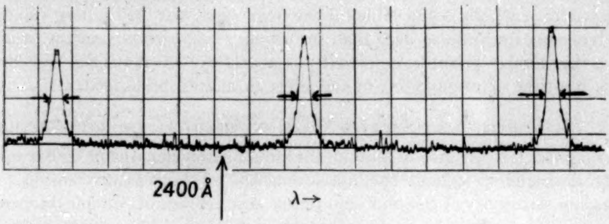


Fig. 7.

The indirect measurements just described, together with some subsidiary measurements and estimates of intensities, have led to the following result. It is possible to resolve at least 1/18 of one order at wavelengths of 2400 Å, and probably considerably less, with transmission factors not substantially below those found at longer wavelengths, but the aluminium films have to be made so thick that they are almost opaque to visible light.

ACKNOWLEDGMENTS

The authors wish to thank the City of Portsmouth Education Department for a grant to one of them (J. C. B.), and St. Hugh's College, Oxford, for a scholarship awarded to another (A. P. P.).

REFERENCES

CABRERA, N., 1944, *C. R. Acad. Sci., Paris,* **218,** 994.

CRAWFORD, M. F., GRAY, W. M., SCHAWLOW, A. L., and KELLY, F. M., 1949, *J. Opt. Soc. Amer.*, **39,** 888.

KUHN, H., and WILSON, B. A., 1950, *Proc. Phys. Soc.* B, **63,** 745.

SENNETT, R. S., and SCOTT, G. D., 1950, *J. Opt. Soc. Amer.*, **40,** 203.

The Determination of Infra-Red Absorption Spectra from Reflection Measurements

By T. S. ROBINSON AND W. C. PRICE
Wheatstone Laboratory, King's College, London

Communicated by J. T. Randall; *MS. received* 11*th May* 1953

Abstract. The theory of infra-red reflection spectra has been applied (i) to obtain the frequencies and extinction coefficients of the intense absorption bands in polytetrafluorethylene near 8·5 microns, (ii) to determine the number and states of polarization of the fundamental NH stretching absorption bands in a crystal of urea. The reflection method has definite advantages over transmission measurements in both cases.

---

§ 1. INTRODUCTION

THE study of the infra-red absorption spectra of crystals and oriented polymers with polarized radiation has recently been applied with considerable success in the elucidation of their structures. Such investigations have been made, for example, by Ambrose, Elliott and Temple (1949, 1951), by Halverson and Francel (1949) and by Newman and Badger (1951). One of the most serious limitations in the development of the technique has been the difficulty in the preparation of suitable samples (Newman and Halford 1950). For strong absorption bands the optimum sample thickness may be less than one micron yet the sample area should cover the spectrometer entrance slit. Fragments of material of such dimensions are often difficult to produce and handle. If the substance to be examined is crystalline several thin sections may be required each cut normally to different crystallographic directions. Some relief from this difficulty is obtained with the aid of an infra-red microscope of sufficient numerical aperture. The length and breadth of the sample may then each be reduced by about a factor of ten. It should be noted, however, that when polarized radiation is used a high N.A. may lead to spurious estimates of dichroism (Newman 1952).

In view of this obstacle to transmission measurements it was considered worth while to investigate the possibility of obtaining absorption spectra from measurements of reflectivity. While surface films and surface scatter affect considerably measurements of reflectivity in the visible, in the infra-red beyond $2{\cdot}5\,\mu$ their effects can be neglected. For observations of reflected radiation the sample need have only one flat surface of sufficient area. The optical system must discriminate against radiation other than that reflected from this surface. This is achieved by imaging the reflecting surface on to the spectrometer entrance slit and by choosing a sufficiently thick sample, so that reflections from back surfaces are negligible.

§ 2. DETERMINATION OF ABSORPTION SPECTRA BY REFLECTION

Several methods exist for the determination of the optical constants of a substance from reflectivity data at any fixed wavelength (Avery 1952, Simon 1951, Simon and McMahon 1953). They all depend essentially upon oblique incidence measurements and are therefore not directly applicable to anisotropic substances. Simon (1951) has applied the impedance concept to the normal incidence reflection spectrum and from it he has shown how the absorption spectrum may be deduced. In his treatment all absorption bands are assumed to have the same shape. These matters were discussed in a recent note (Robinson 1952) where it was indicated how the absorption spectrum could be derived from the general case of a normal incidence reflection spectrum. The reflected wave was regarded as the attenuated output from an equivalent electric network excited by the incident wave. From established principles, the phase change on reflection was deduced from the curve of reflection attenuation against frequency. The optical constants $\mathbf{n}=n-ik$ were then determined with the aid of a Smith chart. The substance studied was polythene in the 3 μ region. The same method has been applied with special advantage in the following two cases. In the first case the absorption is so intense that it is impracticable to cut samples thin enough to measure either the positions of the band maxima or their absorption coefficients by transmission methods. In the second case the need for growing extremely thin crystals has been avoided.

§ 3. ABSORPTION SPECTRUM OF POLYTETRAFLUORETHYLENE IN THE 8·5 μ REGION

The only published absorption spectrum of polytetrafluorethylene (P.T.F.E.) is one which shows complete absorption in the 8·5 μ region for a film of the material 51 μ thick (Smith 1949). It was therefore decided to investigate this region with the reflection equipment. The reflection attenuation spectrum from 1000 to 1600 cm^{-1} is shown in fig. 1. By the method already described (Robinson 1952) the magnitude and phase of the complex reflection coefficient $\mathbf{r}$ were determined and plotted on a Smith chart, fig. 2. Extinction coefficient values have been read off from this chart and the resulting absorption spectrum, after application of the correction described in § 5, is shown in fig. 3.

Two absorption bands of approximately equal peak intensity occur at 1154 cm^{-1} and 1213 cm^{-1} respectively. These two bands are probably associated with C–F bond stretching vibrations. Confirmation of this assignment was obtained from observations of dichroism for the two absorption bands. By hot rolling at 125°c it was found possible to produce oriented films of P.T.F.E. a few microns thick. The absorption near 8·5 μ in unoriented films of this thickness would be too great for reliable estimates of extinction coefficients. With the oriented films however there was no doubt that the 1154 and 1213 cm^{-1} bands showed stronger absorption with the electric vector perpendicular to the direction of rolling, and presumably perpendicular to the molecular chains.

Detailed analyses of the normal modes of vibration of an infinite chain of CH_2 groups, where the carbon atoms are co-planar, have been applied to the observed infra-red spectrum of polythene (Sheppard 1950, Kellner 1951). A similar treatment for P.T.F.E. requires investigation of the remaining absorptions of the material. In this connection the oriented films have been examined for dichroism in the two regions of strong absorption near 16 μ and 19 μ. Four distinct absorption bands were found: the two most intense at 641 and 513 cm^{-1} with parallel dichroism, one of medium strength at 556 cm^{-1} with perpendicular dichroism and the fourth at 630 cm^{-1} appeared as a shoulder on the 641 cm^{-1} band. In the type of infinite CF_2 chain considered only one

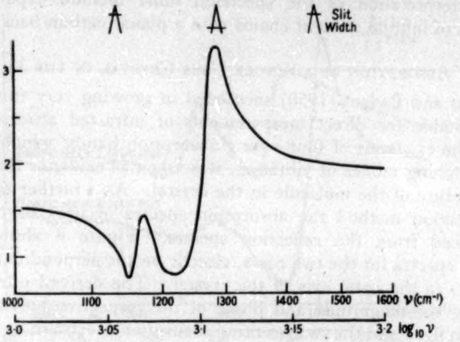


Fig. 1. Reflection attenuation, $-\ln|\mathbf{r_s}|$, of P.T.F.E. between 1000 and 1600 cm^{-1}.

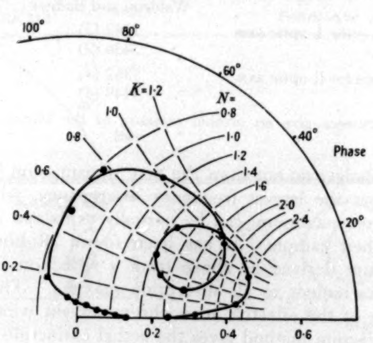


Fig. 2. Amplitude and phase of complex reflection coefficient of P.T.F.E. between 1100 and 1350 cm^{-1}. Superimposed Smith chart gives values of extinction coefficient K and refractive index N.

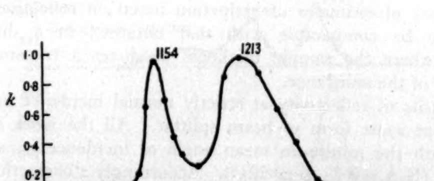


Fig. 3. Absorption spectrum showing extinction coefficient k of P.T.F.E. as derived from reflection spectrum.

very strong parallel band, of symmetry type B_{1u}, is to be expected. It is clear that an interpretation of the spectrum must include departures from the assumption of infinite straight chains with a planar carbon backbone.

§4. ABSORPTION OF A SINGLE UREA CRYSTAL IN THE 3 μ REGION

Waldron and Badger (1950) succeeded in growing very thin single crystals of urea suitable for direct measurements of infra-red absorption in the 3 μ region. The existence of four strong absorption bands, which they attributed to NH stretching modes of vibration, was taken as evidence for the completely planar structure of the molecule in the crystal. As a further test of the power of the reflection method the absorption spectra of urea in this region have been deduced from the reflection spectra. Figure 4 shows the reflection attenuation spectra for the two cases, electric vector perpendicular and parallel, respectively, to the optic axis of the crystal. The derived plots on the Smith chart giving the magnitude and phase of the respective reflection coefficients are shown in fig. 5 and the two resulting absorption spectra in fig. 6. Comparison of these spectra with those of Waldron and Badger reveals a general agreement. The frequencies (in cm^{-1}) of the four intense absorption maxima are :

	Waldron and Badger	This work
Electric vector ⊥ optic axis	3342 (3)	3346 (4)
	3436 (2)	3436 (2)
Electric vector ‖ optic axis	3362 (4)	3353 (3)
	3449 (1)	3460 (1)

Figures in parentheses give an ordinal measure of the intensities of the absorption maxima.

Waldron and Badger do not show the zero transmission line in their spectra and it may be that the largest frequency discrepancy, which occurs for the strongest absorption maximum, can be partially explained in terms of the low transmission of their sample. It has been found (Robinson 1953) that an absorption spectrum derived as above from a reflection spectrum imparts a slight frequency increment to all absorption maxima. The shift amounts to about one-quarter of the effective monochromator slit width. It will also be noted that the reflection method gives the actual extinction coefficient whereas in a direct absorption measurement only relative values are obtained unless the sample thickness is known.

§5. GENERAL OBSERVATIONS ON THE REFLECTION METHOD

The accuracy of estimates of extinction based on reflection measurements is expected to be comparable with that obtained in a direct absorption measurement where the sample thickness produces a transmission equal to the reflectivity of the substance.

Measurements of reflectivity at strictly normal incidence are naturally not possible without some form of beam splitter. All the work so far has been carried out with the minimum mean angle of incidence consistent with the beam aperture (N.A. $=0{\cdot}3$; $\phi \simeq 22\frac{1}{2}^{\circ}$). Accordingly a correction for obliquity, obtained as follows, has been incorporated in the derivation of the above results. At normal incidence the complex reflection coefficients are given by

$$\mathbf{r}_s=\mathbf{r}_p=\mathbf{r}=\frac{\mathbf{n}-1}{\mathbf{n}+1} \qquad \ldots\ldots (1)$$

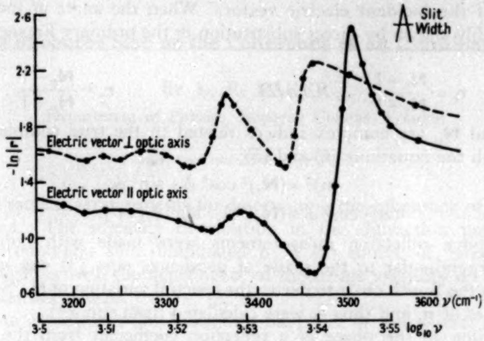


Fig. 4. Reflection attenuation spectra of single crystal of urea in 3 μ region.

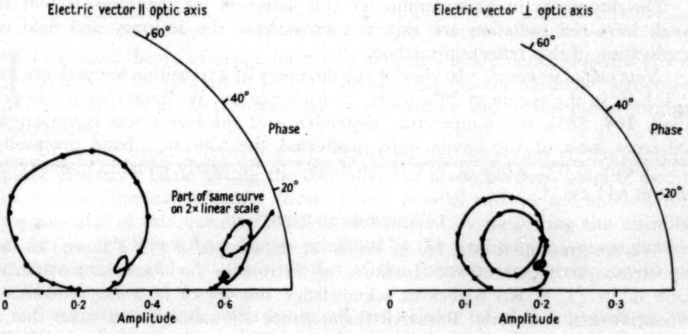


Fig. 5. Amplitude and phase of complex reflection coefficients of single crystal of urea in 3 μ region.

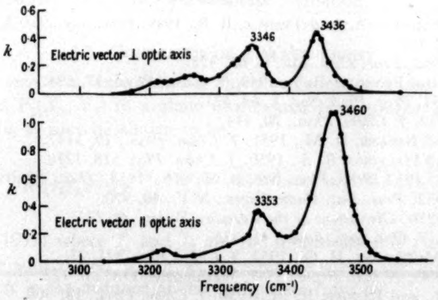


Fig. 6. Absorption spectra of single crystal of urea as derived from reflection spectra.

where $\mathbf{n} = n - ik$ = complex index of refraction and ' s ' and ' p ' indicate the orientation of the incident electric vector. When the angle of incidence is ϕ it can be readily shown by direct substitution in the ordinary Fresnel equations, that

$$\mathbf{r}_s = \frac{\mathbf{N}_s - 1}{\mathbf{N}_s + 1} \quad \ldots\ldots (2) \qquad \mathbf{r}_p = \frac{\mathbf{N}_p - 1}{\mathbf{N}_p + 1} \quad \ldots\ldots (3)$$

where $\mathbf{N}_s$ and $\mathbf{N}_p$ are complex indices related to the true complex refractive index through the equations (4) and (5) :

$$(\mathbf{n})^2 = (\mathbf{N}_s)^2 \cos^2 \phi + \sin^2 \phi, \quad \ldots\ldots (4)$$

$$(\mathbf{n})^2 \cos^2 \phi = (\mathbf{N}_p)^2 \{1 - \sin^2 \phi / (\mathbf{n})^2\}. \quad \ldots\ldots (5)$$

Since the above reflection measurements were made with radiation plane polarized perpendicular to the plane of incidence, eqn. (3) was used in conjunction with the Smith chart to derive the spectral variation of $\mathbf{N}_s$. Subsequently the true values of $\mathbf{n}$, and thus k, were calculated from eqn. (4).

Computation of the phase of a reflection coefficient from the attenuation spectrum is rather tedious. It appears that some form of analogue computer, as used in electric network calculations, may simplify the procedure (Scott 1952).

Developments in photoconductive cell detectors for the measurement of weak infra-red radiation are expected to increase the accuracy and field of application of this reflection method.

Note added in proof. In view of the discovery of a transition temperature for P.T.F.E. in the region of 20°c by H. A. Rigby and C. W. Bunn (*Nature, Lond.*, 1949, **164**, 583), the temperature dependence of the bands was investigated. Whereas most of the bands were unaffected, the 630 cm^{-1} band practically disappears at 0°c, while at 50°c it was found to be nearly equal in intensity to that at 641 cm^{-1}.

ACKNOWLEDGMENTS

We are grateful to Mr. H. A. Willis for the sample of P.T.F.E. and to the Spectroscopic Panel of the Institute of Petroleum for financial assistance. One of us (T. S. R.) wishes to acknowledge the tenure of a Department of Scientific and Industrial Research maintenance allowance and to state that a part of this work has been abstracted from a Thesis submitted for the Ph.D. degree in the University of London.

REFERENCES

AMBROSE, E. J., ELLIOTT, A., and TEMPLE, R. B., 1949, *Proc. Roy. Soc.* A, **199**, 183; 1951, *Ibid.*, **206**, 192.
AVERY, D. G., 1952, *Proc. Phys. Soc.* B, **65**, 425.
HALVERSON, F., and FRANCEL, R. J., 1949, *J. Chem. Phys.*, **17**, 694.
KELLNER, L., 1951, *Proc. Phys. Soc.* A, **64**, 521.
NEWMAN, R., 1952, *J. Chem. Phys.*, **20**, 444.
NEWMAN, R., and BADGER, R. M., 1951, *J. Chem. Phys.*, **19**, 1147.
NEWMAN, R., and HALFORD, R. S., 1950, *J. Chem. Phys.*, **18**, 1276.
ROBINSON, T. S., 1952, *Proc. Phys. Soc.* B, **65**, 910; 1953, *Thesis*, University of London.
SCOTT, R. E., 1952, *Proc. Inst. Radio Engrs., N.Y.*, **40**, 970.
SHEPPARD, N., 1950, *Discussions of the Faraday Society*, **9**, 272.
SIMON, I., 1951, *J. Opt. Soc. Amer.*, **41**, 336.
SIMON, I., and McMAHON, H. O., 1953, *J. Chem. Phys.*, **21**, 23.
SMITH, D. C., 1949, *N.R.L. Report* 3567, Washington, D.C., p. 172.
WALDRON, R. D., and BADGER, R. M., 1950, *J. Chem. Phys.*, **18**, 566.

The Effect of Source Size on the Coherence of an Illuminating Wave

By L. R. BAKER
Department of Physics, Imperial College, London

MS. received 25th June 1953

Abstract. Practical arrangements for determining the coherence of a wave front are discussed. The intensity distribution in the diffraction pattern of two parallel interferometer slits illuminated by a slit source of variable width is calculated together with the dependence of the fringe visibility on the choice of interferometer slit widths.

Finally an experiment demonstrating the effect of source size on coherence is described and the results are compared with those to be expected from the above analysis.

---

§ 1. Introduction

In a general theory of image formation in optical instruments Hopkins (1953) has utilized the concept of partial coherence for the elucidation of structures observed both in the classical type of microscope and in the phase contrast instrument.

An earlier paper (Hopkins 1951) explained the derivation of the complex phase coherence factor giving the correlation in phase between pairs of points on a plane illuminated from a distant source parallel with it, and showed that the modulus of this quantity could be determined by measuring the visibility of the Young's interference fringes produced by the secondary waves originating at these points, and that the argument of the function was merely concerned with the location of the fringes.

Some experimental confirmation of the validity of this approach to the problem of image formation was clearly desirable. The particular case considered here represents the simplest illustration of the effect of coherence on an image merely by isolating localized regions of the wave front which subsequently interfere in the process of image formation.

§ 2. Coherence of a Wave Front

The coherence of a wave front or the correlation in phase between points $P_1(X_1, Y_1)$ and $P_2(X_2, Y_2)$ in a plane illuminated by a source Σ with intensity distribution $I(u, v)$ may be shown to be

$$\gamma_{12}=\frac{1}{R^2(I_1I_2)^{1/2}}\int_\Sigma I(u,v)\exp\left[ik\{u(X_1-X_2)+v(Y_1-Y_2)\}\right]d\sigma \qquad \ldots\ldots(1)$$

(see Hopkins 1951) where I_1 and I_2 are the intensities at P_1 and P_2, $d\sigma$ is an element of the source over which we integrate, $k=2\pi/\lambda$ the usual propagation constant, and R is the distance of the two points from $d\sigma$.

We note that this can be expressed, apart from a constant factor, by the Fourier transform of the distribution of intensity in the source. If we isolate these disturbances by means of pinholes and later allow them to recombine the visibility of the interference fringes formed will be a measure of their coherence. The state of coherence of the whole wave front could then be examined by exploring the illuminated area using pinholes of a variable separation.

In order to secure sufficient light transmission for accurate visibility measurements the two disturbances were isolated by slits instead of pinholes. It was found convenient moreover to maintain a constant separation of the interferometer slits and to vary the coherence between them by adjusting the angular subtense of the slit source.

The separation and width of the interferometer slits were carefully selected so as to produce sufficiently intense fringes with the largest period. A compromise was attained by the use of slits whose separation was not too large compared with their width and in this case there could be a degree of correlation in phase not only between the two slits but also over the finite relative area of either. In view of this arrangement it was considered necessary to evaluate the distribution of intensity across the diffraction pattern of two parallel slits of finite relative width illuminated by a source having a variable angular subtense.

§ 3. Theoretical Considerations

The problem with which we are now concerned can be regarded as the determination of the distribution of intensity in the image of an incoherent slit shaped object viewed by an objective whose aperture is restricted to two fine parallel slits.

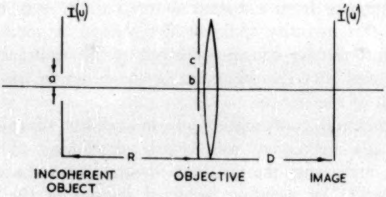


Fig. 1. Schematic representation of the apparatus. $I(u)$ is the slit source whose width can be adjusted so as to vary the coherence in the plane of the interferometer slits.

We shall use in principle the method developed by Duffieux (1946) for the study of incoherent objects. Confining ourselves to one dimension, let the complex transmission of the lens system be $f(x)$ so that the amplitude in the image of a line source at u in the object plane will be

$$F(u'-u)=\frac{1}{(2\pi)^{1/2}}\int_{-\infty}^{+\infty} f(x)\exp\{i(u'-u)x\}dx$$

where u′ is a point in the image plane.

If now the intensity at u′ is denoted by $H(u'-u)$, then

$$H(u'-u)=|F(u'-u)|^2$$

and using this definition we can express the distribution of intensity in the image $I'(u')$, in terms of that in the object $I(u)$ by means of the relation:

$$I'(u') = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} I(u)H(u'-u)\,du$$

the factor $1/(2\pi)^{1/2}$ implying a convenient choice of the unit of intensity.

Using now the Faltung theorem for Fourier transforms (see Titchmarsh 1937), this may be written as

$$I'(u') = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} i(s)h(s) \exp(iu's)\,ds \qquad \ldots\ldots(2)$$

where $i(s)$ is the inverse transform of the object function and $h(s)$ is the inverse transform of $H(u'-u)$ or by definition of the squared modulus of $F(u'-u)$ so that:

$$h(s) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} f(x)f^*(x-s)\,dx.$$

For purposes of symmetry the origin can now be moved to $s/2$ and when normalized so that an image of uniform unit intensity is obtained from such an object we have

$$h(s) = \frac{1}{A} \int_{-\infty}^{+\infty} f(x+s/2)f^*(x-s/2)\,dx \text{ where } A = \int_{-\infty}^{+\infty} |f(x)|^2 dx.$$

Ignoring the aberration we may say that the complex transmission $f(x)$ of the lens system can be taken as unity inside the area covered by the slits and zero outside. $h(s)$ will therefore be equal to the area common to the two aperture systems displaced a distance s relative to each other, as suggested in fig. 2.

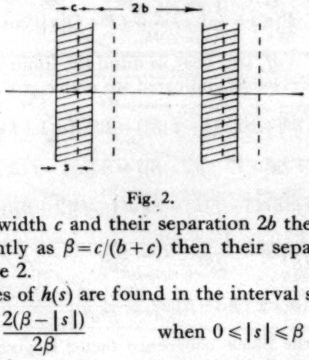


Fig. 2.

If the slits are each of width c and their separation $2b$ their fractional width can be expressed conveniently as $\beta = c/(b+c)$ then their separation will be $2(1-\beta)$ and the overall aperture 2.

The following values of $h(s)$ are found in the interval specified

$$h(s) = \frac{2(\beta - |s|)}{2\beta} \qquad \text{when } 0 \leqslant |s| \leqslant \beta$$

$$h(s) = 0 \qquad \text{when } \beta \leqslant |s| \leqslant 2(1-\beta)$$

$$h(s) = \frac{|s| - 2(1-\beta)}{2\beta} \qquad \text{when } 2(1-\beta) \leqslant |s| \leqslant 2-\beta$$

$$h(s) = \frac{2-|s|}{2\beta} \qquad \text{when } 2-\beta \leqslant |s| \leqslant 2$$

and so we see that $h(s)$ is always of the form $P + Q|s|$. The remaining quantity in eqn. (2) is $i(s)$ and this by definition will be

$$i(s) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{+\infty} I(u) \exp(-ius)\,du$$

and if as in our case the source is of uniform intensity and of semi-width a the object intensity function $I(u)$ will merely serve to limit the range of integration. Thus

$$i(s)=\frac{1}{(2\pi)^{1/2}}\int_{-a}^{+a}\exp(-ius)\,du=\left(\frac{2}{\pi}\right)^{1/2}\frac{\sin as}{s}$$

and eqn. (2) now becomes

$$I'(u')=\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\sin as}{s}(P+Q|s|)\exp(iu's)\,ds.$$

For the purposes of computation it is convenient here to normalize this expression to make $I'=1$, for $u'=s=0$ and moreover to use the intensity per unit size of source in order to make this normalization independent of the size of the source. In this case

$$I'(u')=\frac{\dfrac{1}{a}\displaystyle\int_{-\infty}^{+\infty}\sin as/s(P+Q|s|)\exp(iu's)\,ds}{\displaystyle\int_{-\infty}^{+\infty}P\,ds}$$

where we see that

$$\int_{-\infty}^{+\infty}P\,ds=\int_{0}^{\beta}ds-\int_{2(1-\beta)}^{2-\beta}\frac{(1-\beta)}{\beta}\,ds+\int_{2-\beta}^{2}\frac{1}{\beta}\,ds=2\beta$$

and since $I'(u')$, $i(s)$ and $h(s)$ are even functions it follows that $\exp(iu's)$ must also be even and so consist only of the real part, namely $\cos u's$.

Rewriting then with the range of integration extending over a half plane

$$I'(u')=\frac{1}{\beta}\int_{0}^{\infty}\frac{\sin as}{as}(P+Q|s|)\cos u's\,ds.$$

If now we substitute $u'=af$, and bear in mind the limits appropriate to the three ranges of integration previously indicated we arrive at the following expression:

$$\begin{aligned}I'(u')=\frac{1}{a}\Big\{&A(\mathrm{Si}[a(1+f)\beta]+\mathrm{Si}[a(1-f)\beta])-B(\mathrm{Si}[a(1+f)(2-\beta)]\\&-\mathrm{Si}[a(1+f)2(1-\beta)]+\mathrm{Si}[a(1-f)(2-\beta)]-\mathrm{Si}[a(1-f)2(1-\beta)])+C(\mathrm{Si}[a(1+f)2]\\&-\mathrm{Si}[a(1+f)(2-\beta)]+\mathrm{Si}[a(1-f)2]-\mathrm{Si}[a(1-f)(2-\beta)])\Big\}-\sum_{+-}D\end{aligned}$$

where $$A=\frac{1}{2\beta},\quad B=\frac{1-\beta}{2\beta^2},\quad C=\frac{1}{2\beta^2}$$

and $$D=2\cos^2[a(1\pm f)(2-\beta)/2]\,\frac{\sin^2[a(1\pm f)\beta/2]}{\beta^2a^2(1\pm f)}.$$

It has been seen that the phase coherence factor is given by the transform of the distribution of intensity in the source, and if this is normalized to make $\gamma_{11}=1$ and $\gamma_{12}=1$ when $a\to 0$ in our case $\gamma_{12}=\sin as/as$. The turning points of this function occur when $as=0$, $1{\cdot}43\pi$, $2{\cdot}46\pi$, $3{\cdot}47\pi$, and since $s=2-\beta$ it is clear that the values of a giving maxima of visibility will be

$$a=\frac{1{\cdot}43\pi}{2-\beta},\ \frac{2{\cdot}46\pi}{2-\beta},\ \text{etc.}$$

For the zeroes of visibility or incoherent illumination

$$as=\pi,\ 2\pi,\ 3\pi,\ \text{or}\ a=\frac{\pi}{2-\beta},\ \frac{2\pi}{2-\beta},\ \ldots\ \text{etc.}$$

The above relation for $I'(u')$ is not altogether suitable when a is very small but this case can be simply dealt with by returning to eqn. (2) where $I(u)$ is now a line source of negligible width and so can be represented by a Dirac delta function $\delta(u)$. In this case then $i(s)=1/(2\pi)^{1/2}$ and the normalized value of $I'(u')$ will be

$$I'(u')=\frac{1}{\beta}\int_0^\infty (P+Q|s|)\cos u's\,ds$$

and substituting the limits associated with the various values of P and Q we finally arrive at

$$I'(u')=\left(\frac{\sin\frac{1}{2}u'\beta\cos\frac{1}{2}u'(2-\beta)}{\frac{1}{2}u'\beta}\right)^2$$

which is of course identical to the square of the amplitude in the diffraction pattern deduced in the usual manner when the incident wave is completely coherent.

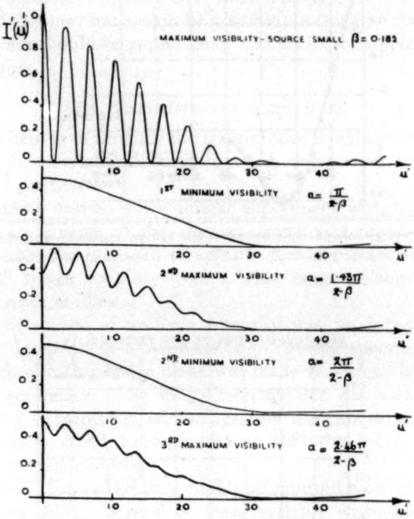


Fig. 3. Plot of the normalized intensity in the image for various source sizes.

For the computation of $I'(u')$ a value of β was taken from the measured dimensions of the actual interferometer slits used in the experiment and this was found to be 0·182.

The distribution of intensity in the diffraction pattern of two parallel slits characterized by β was determined theoretically for the first three maxima and the first two minima of visibility and the results of these calculations are shown in fig. 3.

We see, as indeed was expected, that the interference fringes are always contained within the diffraction envelope of one slit and so for slits whose width is comparable with their separation it is clear that Michelson's definition of visibility no longer holds.

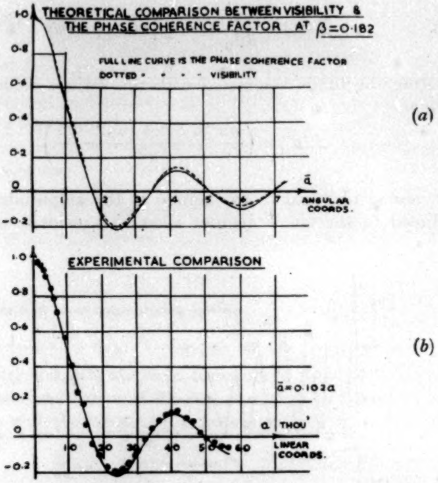


Fig. 4. (*a*) Theoretical comparison between the fringe visibility and the phase coherence factor for a range of source sizes at $\beta=0{\cdot}182$. (*b*) Experimental comparison of fig. 4(*a*).

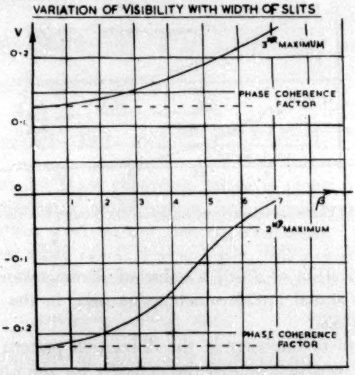


Fig. 5. Variation of visibility with the width of interferometer slits shown for source sizes giving the second and third maximum of visibility.

If I_1, I_3 and I_2, I_4 represent the intensities of the first two maxima and minima we shall redefine the visibility as

$$V=\frac{I_1+I_3-(I_2+I_4)}{I_1+I_2+I_3+I_4}.$$

The variation of the visibility of the fringes with source size when the interferometer slits have a width characteristic $\beta=0{\cdot}182$ was computed and this is compared with the curve obtained when the slits are of negligible width and their separation has been adjusted to give a disappearance of the fringes at the same source size. This comparison is to be seen at the top of fig. 4.

From the previous analysis it is now quite a simple matter to determine the difference between the modulus of the phase coherence factor and the visibility of the fringes as defined above and to see how the latter changes with the width of the slits. This result shown for the second and third maxima of visibility in fig. 5 will be of interest whenever one is concerned with measuring the phase coherence factor, or indirectly with determining the distribution of intensity in a source by the method of interference, since although one would prefer to use small values of β this may not always be possible; in this case the errors involved by this compromise should be appreciated, and made not greater than the other experimental errors.

§ 4. EXPERIMENTAL PROCEDURE

Our concern here is to produce the interference fringes previously described and to see how their variation of visibility with angular source size agrees with that theoretically predicted.

Michelson (1890) carried out a similar experiment with the aim of making very precise measurements of the angular source size required to produce a disappearance of the fringes. There appears to be no account however of the determination of fringe visibility over a wide range of source sizes for the purpose of coherence studies.

An incoherent source of uniform intensity was produced by focusing a mercury vapour lamp on to a symmetrically opening spectrometer slit. The region of coherence in this image plane was made very small by the use of high aperture condenser lenses. The 4358 Å mercury line was isolated by a gelatine filter which has a transmission of 14% at this wavelength and 2% at 4078 Å. This latter impurity could be neglected so that the source was sufficiently monochromatic for our purpose.

Two fine parallel lines ruled on a doubly aluminized piece of glass constituted the interferometer slits. Several of these systems were made, and using a microscope, the best was selected for equality of width, sharpness of the edges, and accuracy of parallelism. The slits were made parallel to the slit source by adjusting for maximum visibility.

The visibility of the fringes was measured photographically using an Ilford ordinary fine grain plate.

The correct exposure was that giving fringes of the maximum visibility for a given source width, since under these conditions we are sure of operating on the linear part of the characteristic curve.

The plate was calibrated after the main experiment by three-quarters closing the plate holder so as just to accommodate a sensitometer wedge in the remaining quarter. The plate was then exposed through the wedge in a plane conjugate with the slit source brought into being by removing the interferometer slits and focusing with the aid of an auxiliary lens.

A microdensitometer was used to relate any fringe to a certain line on the exposed wedge area having the same density and to relate this in turn to the density at the corresponding point of the sensitometer wedge. The visibility of a set of fringes in terms of the maximum and minimum exposures was then calculated from the anti-logs of these corrected densities.

Care was taken in setting up the microdensitometer to ensure that the emulsion itself was at right angles to the optical axis otherwise it could move out of focus with lateral shifts of the plate. The photometer slit width was the largest consistent with no reduction in the fringe visibility, and its total length was utilized.

For small widths of the source producing clearly visible fringes only the maximum and minimum galvanometer deflections were recorded. After the first disappearance however it was found necessary to plot the galvanometer deflections against small lateral shifts of the plate. By drawing the best curve through these points the effect of dust on the plate and local irregularities in the emulsion could be nullified.

For comparison with this experimental curve the phase coherence factor, or visibility of the fringes using interferometer slits of negligible relative width, was found from the first equation where the source is now a slit of width *a* and is of uniform intensity. Thus

$$\gamma_{12} = \sin \frac{kab}{R} \Big/ \frac{kab}{R}.$$

The lower part of fig. 4 indicates the variation of this modulus with *a* compared with the best experimental curve normalized here to overcome the effect of light scattered in the apparatus.

§ 5. CONCLUSION

When careful attention was paid to the details of visibility measurement previously outlined, the errors recorded over the range of source sizes before the first disappearance of the fringes were as low as 2%. For the larger source sizes however, where the visibility of the fringes was now considerably reduced, the errors rose to 5% and were unavoidable due to the technique of photographic photometry employed.

We note that these errors are of the same order as the 6% difference between the modulus of the phase coherence factor and the visibility expected at second maximum from the theoretical results. It was found possible by means of a repetition of the experiment to select mean values of the visibility curve at these large source sizes and fig. 4 shows that these do lie at least on the correct side of the phase coherence curve, and have approximately the deviation that we might expect.

The possibility of determining the correlation in phase between two disturbances in a wave front from a measurement of the visibility of the fringes formed when they interact has been demonstrated.

The slight deviations observed between these quantities are well within those to be expected from a practical realization of this theoretical equality.

ACKNOWLEDGMENTS

I wish to thank Dr. H. H. Hopkins for suggesting this problem and for helpful discussion concerning it. Thanks are also due to the Department of Scientific and Industrial Research for the provision of a maintenance grant.

REFERENCES

DUFFIEUX, P. M., 1946, *L'integral de Fourier et les application a l'optique* (Besançons: privately printed).
HOPKINS, H. H., 1951, *Proc. Roy. Soc.* A, **208,** 263 ; 1953, *Ibid.*, **217,** 408.
MICHELSON, A. A., 1890, *Phil. Mag.*, **30,** 1.
TITCHMARSH, E. C., 1937, *Theory of Fourier Integrals* (Oxford : University Press).

RESEARCH NOTES

A Method of Estimating Impurity Concentrations in Germanium

By F. W. G. Rose and E. W. Timmins
British Thomson-Houston Research Laboratory, Rugby

Communicated by L. J. Davies; MS. received 10th June 1953

The resistivity ρ of single crystal samples of germanium has been measured in the temperature range $T = 100$–400°K using both d.c. and a.c. (3 kc/s) methods. Since Hall coefficient measurements on the same samples showed a free carrier concentration of the order 10^{13}–$10^{14}\,\mathrm{cm}^{-3}$, it can be assumed that the electron or hole gas is non-degenerate in this temperature range, so that Boltzmann statistics may be applied.

The general expression for conductivity is $\sigma = \boldsymbol{e}(n\mu_n + p\mu_p) = 1/\rho$, where $\boldsymbol{e}$ is the electronic charge, n and p are electron and hole concentrations, μ_n and μ_p are the respective mobilities. The mobilities are temperature dependent terms, whose values are determined by the amount of lattice and impurity scattering present (Conwell and Weisskopf 1946). An assumption was made that in the extrinsic conductivity range of temperatures the carrier and impurity concentrations were the same. If these are of the order 10^{13} to $10^{14}\,\mathrm{cm}^{-3}$, lattice scattering is predominant. Theoretical curves have been calculated of $\log \rho$ plotted against $1/T$ for lattice scattering only. (The electron rest mass $\boldsymbol{m}$—considered independent of temperature—has been used for these calculations and the hole mass has been taken as $1{\cdot}2\,\boldsymbol{m}$.) On comparison with experiment (fig. 1), agreement was found in some cases, but there were others where the discrepancy was considerable.

One way in which this discrepancy can be explained is by taking impurity scattering into consideration. Thus the assumption of identity between carrier and impurity concentrations is invalid, and an impurity concentration outnumbering the carriers must exist.

As Hall coefficient measurements showed a carrier concentration independent of temperature in the extrinsic range, all donor or acceptor impurities must be ionized. Thus the impurity concentration N_{I} can exceed carrier concentration only if there are donors and acceptors present simultaneously. This conception has been used by Hung and Johnson (1950), but with samples of low resistivity such that the carrier concentration was of the order $10^{15}\,\mathrm{cm}^{-3}$ in the extrinsic range above 100°K. Thus the existence of minority impurities less than $10^{15}\,\mathrm{cm}^{-3}$ could be detected only by measurements at very low temperatures where most of the excess donors or acceptors are no longer ionized.

For electrical neutrality $(n + N_{\mathrm{A}}) - (p + N_{\mathrm{D}}) = 0$, i.e. $n - p = N_{\mathrm{D}} - N_{\mathrm{A}}$, where N_{D} and N_{A} are concentrations of donors and acceptors.

As the discrepancy occurred in the extrinsic range where $n \gg p$, or $p \gg n$, then $n \simeq N_{\mathrm{D}} - N_{\mathrm{A}}$ or $p \simeq N_{\mathrm{A}} - N_{\mathrm{D}}$ and so $N_{\mathrm{I}} = N_{\mathrm{D}} + N_{\mathrm{A}}$ can be much greater than n or p.

Theoretical curves have been calculated of $\log \rho$ plotted against $1/T$ with various values of $N_D - N_A$ and $N_A - N_D$ for a series of values of N_I from 10^{14} to $10^{18}\,\mathrm{cm}^{-3}$, beyond which limit Boltzmann statistics become inapplicable. One set of such curves is shown in fig. 2. In calculating the effective mobility μ in

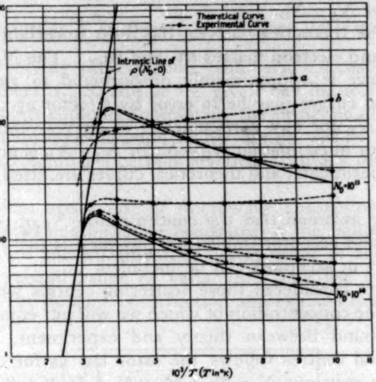


Fig. 1. Comparison of two theoretical curves of log ρ plotted against $1/T$ ($N_D=10^{13}$ and $10^{14}\,\mathrm{cm}^{-3}$) for lattice scattering only, with some experimental results which differ from theory.

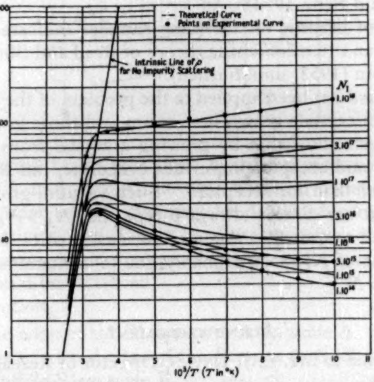


Fig. 2. A set of theoretical curves of log ρ plotted against $1/T$ for $N_D - N_A = 10^{14}\,\mathrm{cm}^{-3}$ with both lattice and impurity scattering in germanium. The appropriate values of N_I are indicated.

the presence of impurity and lattice scattering, a graph (Conwell 1952) was used which enables a correct computation of μ to be made.

The theoretical curves of fig. 2 show that there is quite good agreement with experiment in the extrinsic range. As shown in fig. 2, theory indicates that with increasing values of N_I, the resistivity increases not only in the extrinsic but also in the intrinsic range at any given temperature. Some experimental curves such as curves *a* and *b* in fig. 1 appear to give confirmation to this.

By selecting from the set of theoretical curves that curve which gives the best fit with the experimental curve, one is able to specify the total impurity content N_I and N_D-N_A or N_A-N_D.

In obtaining the theoretical curves, the most uncertain factors seem to be the effective hole and electron masses $\boldsymbol{m}_h$ and $\boldsymbol{m}_e$. The N_I value attributed to any particular curve is approximately proportional to $\boldsymbol{m}^{-1/2}$. Thus the N_I labelling of a set of curves may be in error by a factor up to 3. The value of N_D-N_A or N_A-N_D for any particular set of curves remains practically unaffected. The error in estimating N_D-N_A or N_A-N_D is given by the accuracy with which the experimental and theoretical curves are fitted, and may be within 10%.

So far we have assumed that the concentration N_I of scattering centres is the sum of donor and acceptor impurities, but such lattice imperfections which might act as donors or acceptors can be included.

We have not yet considered those scattering centres which remain neutral (e.g. dislocations) the concentration of which we will put equal to N_0. With the good agreement found between theory and experiment, either $N_0>0$, and scattering by neutral centres follows the same law as for charged centres, or $N_0=0$. Hence we may put $N_I=N_A+N_D+N_0$. If $N_0>0$, but the scattering law is appreciably different, then we should expect poor agreement. Therefore we can assume that in our samples there is no appreciable concentration of such scattering centres for which a different scattering law applies (e.g. lattice dislocations, Dexter and Seitz 1952).

Measurements of lifetime on the same samples indicate a correlation with N_I as suggested by an extension of the theory of Read and Shockley (1952) made by Rose and Ransom (1953, unpublished).

The above method has been applied to the problem of the difference in purity between germanium crystals grown in vacuum and those grown in an inert gas such as argon. Tests show that by growing a crystal from n-type material in vacuum, the donor and acceptor impurities evaporate, but the evaporation rate for donors is greater than for acceptors. Such vacuum-grown crystals can be extremely pure, but are almost always p-type, i.e. $N_A>N_D$. These crystals, on being melted and re-grown in argon of various purities, in all cases show an appreciable increase of N_I most probably due to contamination from the gas, and in general $N_D>N_A$.

ACKNOWLEDGMENT

Our thanks are due to Mr. L. J. Davies, Director of Research, for permission to publish this work.

REFERENCES

CONWELL, E., 1952, *Proc. Inst. Radio Engrs., N.Y.*, **40,** 1327.
CONWELL, E., and WEISSKOPF, V. F., 1946, *Phys. Rev.*, **69,** 258.
DEXTER, D. L., and SEITZ, F., 1952, *Phys. Rev.*, **86**, 964.
HUNG, C. S., and JOHNSON, V. H., 1950, *Phys. Rev.*, **79,** 535.
READ, W. T., and SHOCKLEY, W., 1952, *Phys. Rev.*, **87,** 835.

A Microwave Computing Problem

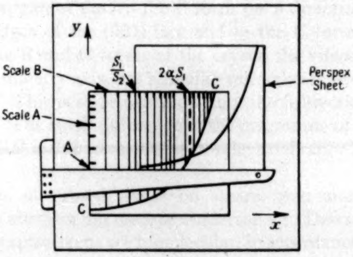
By R. N. BRACEWELL

Division of Radiophysics, Commonwealth Scientific and Industrial Research Organization,
Chippendale, N.S.W., Australia

Communicated by E. G. Bowen; MS. received 31st August 1953

The need for solving the equation $\cosh 2\alpha S_1 = A \cosh 2\alpha S_2$ for α when the real numbers S_1, S_2 and A are given has led to the construction of a special analogue computer employing the principle of the Kelvin bridge (Little 1953). A main consideration of his design was to achieve speed and simplicity whilst maintaining a certain practical accuracy over a desired range of the variables.

Since the problem arises out of microwave absorption measurements, a field of wide interest, the alternative computer of the figure may be of value. It is of simple design and of similar or better order of accuracy for a given size of master curve and range of variables. The use of the computer is described first, and the detailed construction and explanation follow.



Fixed to a drawing board is a sheet bearing a master curve CC and two permanent scales A and B. A movable sheet of Perspex bears the same curve CC on its under surface and may be displaced parallel to itself with a T-square. Alternatively, the T-square, drawing board and Perspex may be dispensed with, leaving a piece of graph paper and a piece of tracing paper, but some speed would be sacrificed.

To obtain a solution,

(i) Set the edge of the T-square to A on scale A.
(ii) Set the left-hand edge of the Perspex to S_1/S_2 on scale B.
(iii) Read off $2\alpha S_1$ on scale B.
(iv) Divide $2\alpha S_1$ by $2S_1$ to get α.

The formulae for the master curve and the two scales may be expressed in terms of rectangular coordinates x and y as follows:

Master curve: $y = \log \cosh e^x$
Scale A: $y = \log A$
Scale B: $x = \log 2\alpha S_1$.

The scale of $2\alpha S_1$ is also used for reading off S_1/S_2.

The procedure may be explained as follows. It may be verified that the desired solution is given by the intersection of the two curves

$$y = \log \cosh e^{x},$$
$$y - \log A = \log \cosh \exp \{x - \log (S_1/S_2)\},$$

where $x = \log 2\alpha S_1$. Now these two curves are the same shape, differing only by a vertical displacement $\log A$ and a horizontal displacement $\log(S_1/S_2)$. The finished computer is simply the embodiment of this result in a mechanical form with convenient provision for performing the displacements and with scales reading directly in A, $2\alpha S_1$ and S_1/S_2.

REFERENCE

LITTLE, V. I., 1953, *Proc. Phys. Soc.* B, **66**, 185.

LETTERS TO THE EDITOR

A Further Note upon the Growth and Optical Properties of Stearic Acid Crystals

In a recent paper (Verma and Reynolds 1953) two types of stearic acid crystals were reported which differed in their optical properties and their edge (or profile) angle. It is possible to identify these two types as the B and C polymorphs (after Piper *et al.* 1926) as a result of the determinations of the *a* and *b* axial dimensions by Schoon (1938) and Anderson and Dawson (1953). The reported optical properties can also be clarified.

For the B polymorph Anderson and Dawson give $a = 5{\cdot}64$ Å, $b = 7{\cdot}59$ Å, whilst for the C form $a = 9{\cdot}11$ Å and $b = 4{\cdot}82$ Å. The angle between the [110] and [1$\bar{1}$0] directions is $2 \tan (b/a)$ which for the B form of the acid leads to a profile angle (acute) $73^\circ 14'$ and for the C form $55^\circ 45'$. These profile angles agree, within experimental error, with the values reported by us, and also with the observations of Müller (1927), Thibaud and Dupré la Tour (1930) and Schoon (1938).

It then becomes apparent that in the B form the *b* direction bisects the acute angle between the edges of the (001) face and in the C form the obtuse angle. Therefore in both the B and C forms of the crystal the vibration parallel to the *b* axis, which is the symmetry axis, has a smaller refractive index than the vibration parallel to the *a* axis. This is to be expected since the *b* direction is perpendicular to the chain-length. The small difference in the magnitude of the birefringences, namely 0·018 and 0·015, is due to the change in the axial ratios and the monoclinic angle.

From their study of growth steps on stearic acid and other long-chain compounds using the electron microscope Anderson and Dawson (1953) conclude that in stearic acid the spiral steps are bimolecular, in accordance with the nature of this acid in solution. Using an optical microscope we have reported growth steps which are integral multiples of the bimolecular unit. These may be considered to be due to a group of bimolecular steps which were not resolved. However, steps which were odd half-integral multiples of the bimolecular unit, i.e. $(65{\cdot}2 \pm 0{\cdot}9)$ Å $= 1{\cdot}5(43{\cdot}5 \pm 0{\cdot}6)$ Å, were also found. These cannot be interpreted as any grouping of unit steps. Further step-height measurements are being undertaken to clarify this.

University of London,
Royal Holloway College,
Englefield Green, Surrey.

Ajit Ram Verma.*
P. M. Reynolds.

9th July 1953; in final form 4th Sept. 1953.

Anderson, N. G., and Dawson, I. M., 1953, *Proc. Roy. Soc.* A, **218**, 255.
Müller, A., 1927, *Proc. Roy. Soc.* A, **114**, 542.
Piper, S. H., Malkin, T., and Austin, H. E., 1926, *J. Chem. Soc.*, 2310.
Schoon, T., 1938, *Z. Phys. Chem.*, **139**, 385.
Thibaud, J., and Dupré la Tour, F., 1930, *C. R. Acad. Sci., Paris*, **191**, 200.
Verma, A. R., and Reynolds, P. M., 1953, *Proc. Phys. Soc.* B, **66**, 414.

* I.C.I. Research Fellow.

REVIEWS OF BOOKS

Photography—Its Materials and Processes, by C. B. NEBLETTE, 5th Edn. Pp. vii + 500. (London: Macmillan, 1953.) 47*s.* 6*d.*

The physicist who uses photography as a tool needs a book which gives him sufficient theory to enable him to understand the processes involved, and which shows him how they can be applied in practice. Such a book should also describe the many photographic techniques available to the scientist at the present time, thus telling him what can be done, and how, so that he can start on the right lines when faced with a photographic problem.

In spite of its title, the work under review is not of this kind, for, apart from fundamental photographic theory, it only deals with ordinary pictorial photography and reproduction processes. It is a collection of articles written by fifteen contributors and compiled by Mr. Neblette who has also taken the largest share in the authorship. On the whole, the emphasis of the book is on the theoretical aspects of the techniques described, and practical hints are given only rarely.

As sometimes happens in books of this sort, the amount of previous knowledge demanded from readers fluctuates strongly from chapter to chapter, and the units and technical terms used are often left undefined (though a definition may sometimes be found in a later chapter). In some of the articles, too, the writing is quite exceptionally clumsy, for example: " The term *polarized*, as applied to light, does not have reference to its color nor its intensity, but to another property which, unlike the other two, is invisible and has to do with the way in which a ray of light vibrates " (p. 73). This quotation will also indicate how low a level of attainment in the reader is expected by some of the authors, particularly those responsible for the chapters on the physics of photography. Much greater demands are made in the sections dealing with chemistry, and there are pages and pages of complicated structural formulae which can hardly be understood by the uninitiated.

However, the book is not uniformly bad. It fails in some of its theoretical sections where it cannot compare with Mees's standard work, or with the excellent little volume by James and Higgins (*Fundamentals of Photographic Theory*, London: Chapman and Hall, 1948), but it contains a number of interesting well-written articles which give a clear scientific account of some of the techniques and materials used in photography and reproduction work, and of the problems which arise in their use. Into this group, in particular, belong the chapters on diffusion-transfer reversal processes, and on the reproduction of tone and colour in monochrome. The chapters on colour photography are excellent as a first introduction, and there are other sections which deserve similar praise. It is a pity that these articles should be found in such poor company.

The book contains many references to original papers, and this will make it more useful to research workers who have a reference library within reach. There are 500 pages of text, many illustrations and quite a few misprints. The price seems reasonable for a book of this size, printed in the U.S.A.

A. J. HERZ.

Dislocations and Plastic Flow in Crystals, by A. H. COTTRELL. Pp. ix + 223. (Oxford : Clarendon Press, 1953.) 25*s*.

It must be apparent by now that no one can safely study any property of crystals without paying regard to dislocations. The subject may be approached from many aspects, of which a few may be mentioned here. From the definition of a lattice : a lattice is the array of points at the intersections of three families of equidistant planes—if one of these families is replaced by a helicoid we have the simplest example of a dislocated lattice. From elasticity theory : the more powerful mathematical methods in this theory relate states of stress to mathematical functions, and the consideration of functions which possess singularities naturally introduces the idea of dislocations. From a pursuit of the idea of ' mosaic structure ' : consideration of the boundaries between mosaic units, with the aid of mathematics or bubble models, shows that the atomic misfit in these boundaries cannot be uniform but will concentrate into dislocation lines—whereupon it becomes clear that the dislocation concept is the more fundamental one and mosaic structure a particular arrangement of dislocations. From the theory of crystal growth: which shows that in many circumstances only dislocated crystals are capable of growing. And from the observed fact of crystal slip : it is scarcely thinkable that slip should be uniform and simultaneous over a macroscopic area of slip surface, and the boundaries between areas which have slipped by different amounts are dislocations, so that these are the natural elements of a theory of plastic deformation of crystalline solids.

Cottrell, a physical metallurgist, naturally adopts this last approach, as the title of his book indicates. One could scarcely write a book on dislocations otherwise at the present moment, since more thought has been given to this aspect of the subject than to any other : the fact that the number of firm conclusions reached is not proportionately large makes all the more welcome the guidance of a physical metallurgist whose judgment is so sound. The elasticity theory and basic geometry of the subject are naturally treated at considerable length, being fundamental to any application; and the author is to be congratulated on his simple and accurate exposition of matters which are only too easily made difficult, and abound in opportunities for error. The parts played by dislocations in crystal boundaries, in crystal growth, and in determining etch patterns are introduced for the valuable purpose of showing that dislocations visibly exist : it is not long since they were widely regarded among metallurgists as a mathematical fiction. Hedges and Mitchell's revelation of dislocations in the interior of a crystal (1953, *Phil. Mag.*, **44,** 223) came just too late for inclusion. Dislocations as the sources, sinks, and traps of vacancies, interstitial atoms and impurity atoms are dealt with in their relationship to mechanical behaviour and annealing : it would have been too early to devote much space to the application of these ideas to other branches of the theory of solids, which are obvious but have been little studied as yet.

Since 1949, Cottrell's 50-page article in Chalmers' *Progress in Metal Physics*, Vol. I, has been the best available introduction to the subject of crystal dislocations, and a handy reference for those already introduced. This fuller account is proportionately more welcome. The reader will have nearly everything that matters in the subject brought to his notice, he will find a number of points which could have puzzled him clearly expounded, and he will be told very little that is wrong without a warning that it may be. F. C. F.

CONTENTS OF SECTION A

PAGE

Prof. D. R. Bates and Major G. Griffing. Inelastic Collisions between Heavy Particles—I: Excitation and Ionization of Hydrogen Atoms in Fast Encounters with Protons and with other Hydrogen Atoms 961

Prof. D. R. Bates and Dr. A. Dalgarno. Electron Capture—III: Capture into Excited States in Encounters between Hydrogen Atoms and Fast Protons 972

Mr. A. M. Lane. Studies in Intermediate Coupling: the Energy States of ^{13}C and ^{13}N belonging to the Configuration 1p^{9} 977

Mr. H. N. V. Temperley. A New Theory of Liquid Helium: Further Treatment 995

Prof. H. Messel and Prof. H. S. Green. The General Three-Dimensional Theory of Cascade Processes 1009

Mr. D. A. Tidman, Dr. E. P. George and Dr. A. J. Herz. The Production of Delta-Rays in Nuclear-Research Emulsions 1019

Dr. J. R. Holt and Dr. T. N. Marsham. An Investigation of (d, p) Stripping Reactions—V: Results for some of the Light Elements and Conclusions . 1032

Mr. A. R. Bodmer. Nuclear Scattering of Electrons and Isotope Shift . . 1041

Dr. W. G. Davey. The Elastic Scattering of 1·33 mev and 2·76 mev Gamma Rays by Lead 1059

Research Notes:

Dr. M. E. Pillow. Intensities in the Herzberg System of O_2 1064

Dr. P. Swan. On the Existence of a Bound State of ^{4}H 1066

Dr. E. Lindholm. Ionization and Fragmentation of Molecules by Bombardment with Atomic Ions 1068

Letters to the Editor:

Mr. E. Fawcett. The Surface Resistance of Normal and Superconducting Tin at 36 000 Mc/s 1071

Dr. E. R. Rae and Miss E. M. Bowey. A Scintillation Detector for Neutrons of Intermediate Energy 1073

Dr. C. J. D. Jarvis. The Internal Conversion Electrons Emitted in the Decay of Ionium and Radio-Thorium 1074

Mr. A. L. Mathur and Dr. K. M. Gatha. Jastrow's Nuclear Model for High Energy Electron Scattering 1075

Mr. W. R. G. Kemp, Mr. A. K. Sreedhar and Dr. G. K. White. The Thermal Conductivity of Magnesium at Low Temperatures 1077

Reviews of Books 1079

Contents of Section B 1080

Fig. 2. Fe; oblique ⟨111⟩ orientation (weak)..

Fig. 3. Pd; oblique ⟨211⟩ orientation (weak).

Fig. 4. Cu; oblique ⟨211⟩ orientation (weak).

Fig. 5. Al; oblique ⟨100⟩ orientation.

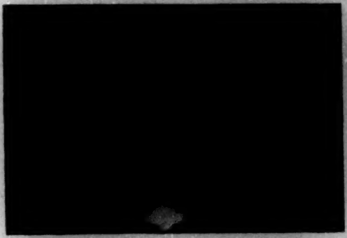


Fig. 6. Al; random (dep. in air at 10^{-3} mm Hg).

Fig. 7. Mg; {103} orientation.

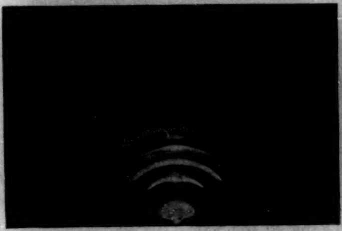


Fig. 8. Bi; {100} orientation (face-centred rhombohedral axes, pseudocubic).

Fig. 9. Zn; {112} orientation, limited azimuthally (+fainter ZnO).

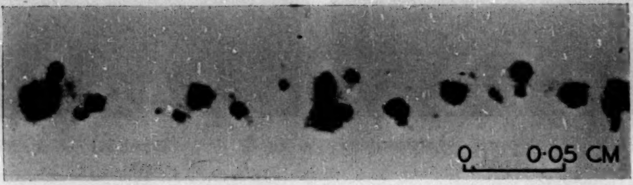


Fig. 1. Autoradiograph of steel surface after passage of clean radioactive copper slider under a load of 2000 g, showing black circular regions caused by copper wear particles.

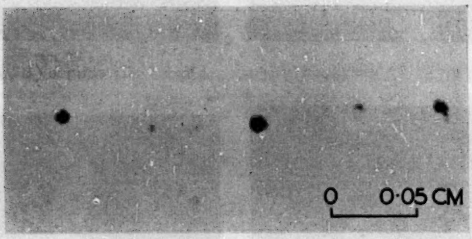


Fig. 8.

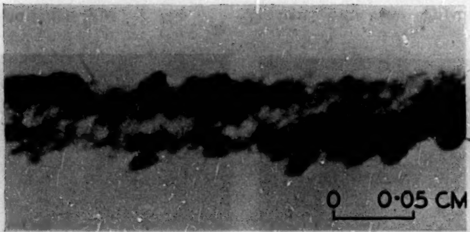


Fig. 9.

Figs. 8 and 9. Autoradiographs of copper surface after passage of radio active steel slider under loads of 50 g (fig. 8) and 5000 g (fig. 9).

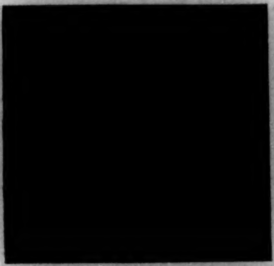


Fig. 2. Thick mica single crystal, 60 kv.

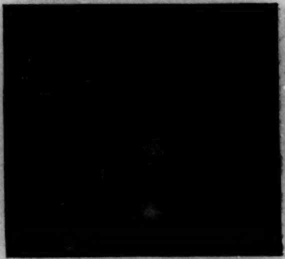


Fig. 3. As fig. 2, but at 100 kv.

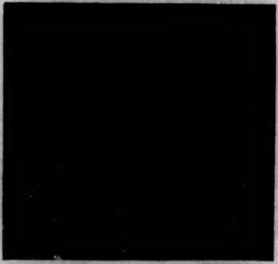


Fig. 4. As fig. 2, but at 140 kv.

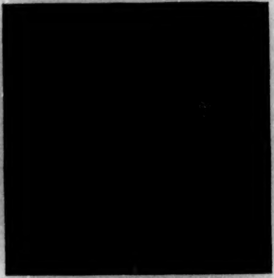


Fig. 5. Rock salt (001) cleavage face, beam || [100], 60 kv.

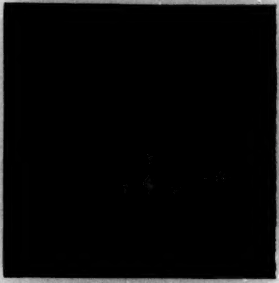


Fig. 6. As fig. 5, but at 120 kv.

PLATE I.

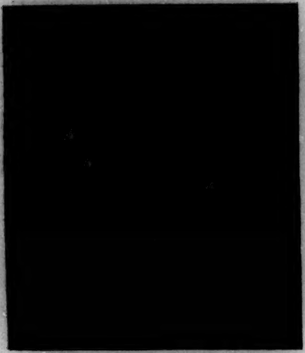


Fig. 7. Quartz single crystal, mechanically polished to a high optical finish; angle of incidence about 1°, 60 kv.

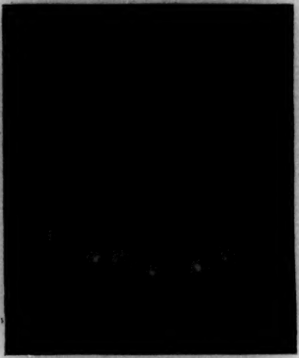


Fig. 8. As fig. 7, but at 140 kv.

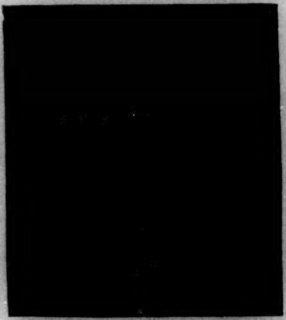


Fig. 9. As fig. 8, but with angle of incidence reduced to a few minutes.

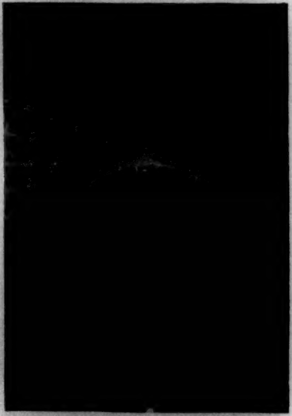


Fig. 11. Rutile single crystal, conchoidal fracture, 120 kv.

Plate II.

Photoelectromagnetic and Photoconductive Effects in Lead Sulphide Single Crystals

By T. S. MOSS
Telecommunications Research Establishment, Great Malvern, Worcs.

MS. received 6th July 1953

Abstract. For the first time bulk photo-effects have been observed in single crystals of lead sulphide. Both photoelectromagnetic and photoconductive measurements have been made, and carrier lifetimes and quantum efficiencies estimated therefrom. Quantum efficiencies are near unity for all samples, while lifetimes are found to range from 6×10^{-10} to 9×10^{-6} second. Measurements on lifetimes by two other methods give values in reasonable accord with the photoelectromagnetic values. Study of the spectral distribution of sensitivity establishes that there is no appreciable change in quantum efficiency with wavelength except for the rapid fall at 3μ, thus establishing that photoelectrically there is no difference between the main absorption band and the 'tail' band. The optical activation energy (i.e. quantum energy at which sensitivity is halved) is 0·41 ev, and it is concluded that the thermal activation energy is very close to this value.

§ 1. Introduction

Although in the past there have been frequent observations of photoconductivity in sensitized layers of PbS, and of photovoltaic and photoconductive effects at metal contacts on single crystals (Gibson 1952), bulk photo-effects in single crystals have not previously been observed.

One reason why photoconductivity has not been observed in the past is no doubt the difficulty of obtaining crystals of the requisite high purity, and another is the difficulty of ensuring that any effect which might have been observed was indeed due to photoconductivity and not to bolometric effects. For the latter case the use of the photoelectromagnetic (PEM) effect gives a powerful method of distinguishing the two by the independent control of the magnetic field.

§ 2. Theory

Two independent theories of the PEM effect have recently been formulated (Moss, Pincherle and Woodward 1953, Aigrain and Bulliard 1953). In the case of the present measurements where the specimens have relatively high carrier concentrations and short diffusion lengths and where the preparation of specimens is more difficult than for germanium, it has been found preferable to adopt a simple phenomenological theory of the short-circuit current in place of the more complex formulae previously derived for voltages set up on the front and back surfaces of the specimen.

Consider a rectangular specimen of length l, thickness t and width w which is irradiated normal to its length and width so that Q quanta per second are absorbed in the material (fig. 1). Photo-electrons and photo-holes generated near the surface will set up a concentration gradient and the carriers will diffuse down through the specimen. Under the influence of a transverse magnetic field (i.e. perpendicular to the length of the specimen and to the direction of the radiation) the two types of carrier will be deflected in opposite directions through the Hall angle $\tan\theta = \mu B$ where μ is the mobility in $m^2 v^{-1}$ sec and B the magnetic induction in weber m^{-2}.

Assuming that the neutrality condition is satisfied, the diffusion length L_e travelled by an electron is given by $L_e = (D_e\tau)^{1/2}$ where D_e is the diffusion constant and τ the average carrier lifetime, provided that there is no appreciable electric field in the direction of motion. In the present problem this means that the diffusion photo-voltage set up in the direction of the thickness does not give a field comparable with $\boldsymbol{k}T/\boldsymbol{e}$ per diffusion length. Estimation of the diffusion voltage using the theory of Moss, Pincherle and Woodward (1953) shows that it is several orders of magnitude less than this. Hence for small magnetic fields the carrier will move a distance $\mu_e B(D_e\tau)^{1/2}$ between the electrodes and give a charge contribution $\boldsymbol{e}\mu_e B(D_e\tau)^{1/2}/l$ to the external circuit. A similar contribution will arise from the photo-holes, so that the net steady short-circuit current generated by the sample will be

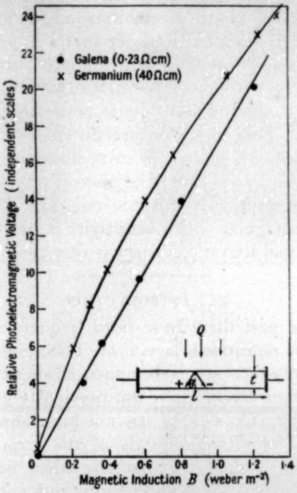


Fig. 1.

$$i_s = \eta \frac{QeB}{l} \{\mu_e(D_e\tau)^{1/2} + \mu_h(D_h\tau)^{1/2}\}^* \qquad \ldots\ldots(1)$$

or

$$i_s = \eta \frac{QeB}{l} \mu^{3/2}(\boldsymbol{k}T\tau/\boldsymbol{e})^{1/2} \text{ since } D_e = \mu_e \boldsymbol{k}T/\boldsymbol{e}$$

* Alternatively, the instantaneous current density j produced by each photo-electron per unit volume moving with velocity v in the direction of the electrodes is $j = \boldsymbol{e}v$ so that the charge transfer is

$$\int_0^\tau wtj\,dt = wt\boldsymbol{e}\int_0^\tau v\,dt = (wt\boldsymbol{e}) \times (\text{distance moved between the electrodes}).$$

Hence the total charge transfer per second due to the electrons gives

$$i_s = \frac{Q\eta}{wlt} wt\boldsymbol{e}\{\mu_e B(D_e\tau)^{1/2}\} = \frac{Q\eta\boldsymbol{e}}{l}\mu_e B(D_e\tau)^{1/2}$$

with a similar term for the hole contribution.

where η is the quantum efficiency and μ is an effective combined mobility for electrons and holes.

It may be noted that the photo-diffusion voltage which is set up in the direction of the irradiation will accelerate the slower carriers (holes) and slow up the faster ones (electrons), and will in fact equalize the path lengths of the two types of carrier—otherwise there would be a net current in the direction of the radiation. For this reason, and also because the individual mobilities are not accurately known in PbS, it is considered preferable to use the above 'effective combined mobility' rather than μ_e and μ_h explicitly.

Equation (1) may be compared with the expression for the short-circuit current derived by Moss, Pincherle and Woodward (1953). For $(D\tau)^{1/2} \ll t$ as is the case for PbS crystals, their expression reduces to

$$i_s = \frac{\eta Q\boldsymbol{e}(\mu_e+\mu_h)\,B(D_h\tau)^{1/2}}{l}\,\frac{\alpha}{1+\alpha}$$

where $\alpha = D_h/Ls$ and s is the surface recombination velocity. For PbS, $D \simeq 25\times10^{-4}$ m^2 sec^{-1} and $L \sim 10^{-5}$ m. The surface recombination velocity s should be small for a cleavage surface—it would be expected to be comparable with the values found for well etched germanium surfaces, i.e. ~ 2 m sec^{-1}. Rough measurements by the method outlined by Moss, Pincherle and Woodward (1953) indicate that $s < 10$ m sec^{-1}. Hence $\alpha \gg 1$ and $\alpha/(1+\alpha) \simeq 1$. The expression is then practically identical with eqn. (1).

The equation for the PEM voltage given by Aigrain and Bulliard (1953) may be converted to short-circuit current by dividing by the specimen resistance (as the behaviour has already been shown to be ohmic in this respect by Moss, Pincherle and Woodward (1953)), and if again unity is ignored with respect to α one obtains

$$i_s = \eta Q\boldsymbol{e}2\mu BL/l \text{ where } \mu = 2\bigg/\left(\frac{1}{\mu_e}+\frac{1}{\mu_h}\right).$$

Thus all three expressions are of the form $i_s = (\eta QB\boldsymbol{e}/l)(\boldsymbol{k}T\tau/\boldsymbol{e})^{1/2}\,\mu^{3/2}$ where μ is a weighted combined mobility, the method of weighting being somewhat different in each case, but all reducing to the same value for $\mu_e = \mu_h$.

For the case of photoconductivity in an applied electric field F v m^{-1} the distance travelled between the electrodes is $\mu_e F\tau$ for an electron and $\mu_h F\tau$ for a hole, so that provided $\mu F\tau$ is small compared with the length of the specimen the current is $i' = \eta Q\boldsymbol{e}\mu' F\tau/l$ where μ' is now the simple sum of $\mu_e + \mu_h$. As the ratio of mobilities is not large for PbS, differences in μ and μ' will be ignored henceforth.

Hence for a given specimen and given illumination,

$$i'/i_s = (F/B)(\boldsymbol{e}\tau/\mu\boldsymbol{k}T)^{1/2} = a \text{ for example,}$$

and

$$\tau = \frac{B^2a^2}{F^2}\,\frac{\mu\boldsymbol{k}T}{\boldsymbol{e}} = \frac{B^2a^2}{F^2}\,D \qquad \dots\dots(2)$$

where D is a weighted total diffusion constant. Thus the mean lifetime of the carriers is determined directly from the ratio of the short-circuit currents under applied electric and magnetic fields.

Similarly

$$\frac{i_s^2}{i'} = \frac{\eta QB^2\boldsymbol{e}}{lF}\,\mu D \text{ or } \eta = \frac{i_s^2}{i'}\,\frac{lF}{QB^2\boldsymbol{e}\mu D}. \qquad \dots\dots(3)$$

From expressions (2) and (3), values of the carrier lifetimes and quantum efficiencies can be obtained within the limits of accuracy imposed by uncertainties in μ and D.

§ 3. EXPERIMENTAL DETAILS

The specimens used were cleavage plates of natural galena from various sources. With care it was possible to obtain samples about 5 mm long, 0·5 mm wide and 0·2–0·3 mm thick. These had the ends copper plated, and fine platinum wires then soldered on.* The electrodes and parts of the crystal in close proximity to them were masked by black paint.

The irradiation was supplied by a double monochromator with a Nernst source, the radiation being chopped mechanically at the entrance to the monochromator. The signal from the specimen was passed through a transformer with an impedance step up of 10^3–10^5 depending on the specimen resistance, and amplified in a high gain amplifier followed by a homodyne detector circuit.† Chopping frequencies of 400–700 c/s were used. The signals obtained were small; at the transformer input they covered the range 10^{-8} to 10^{-6} v. For the smallest signals the bandwidth of the homodyne system was reduced to 1/16 c/s in order to ensure adequate signal/noise ratios.

The amplifier system was calibrated by use of a signal generated by the chopper using a photocell and lamp so that the wave form of the calibrating signal would be the same as the wave form finally measured from the PbS sample. The signal from the photo-cell was measured on a valve voltmeter and then passed through known attenuators to the amplifier input transformer.

The radiant power falling on the specimens was calibrated by focusing the radiation on to a thermopile of known area in the same manner as it was normally focused on to the specimens, and then checking the radiation thermopile against another thermopile of calibrated flux sensitivity in unfocused conditions. A correction was applied for the known reflection loss at the PbS surface.

Measurements of resistivity and electric fields along the specimens (for given working currents) were made by potentiometric methods with tungsten probes operated by a micromanipulator. The electric fields used were such to make a of eqn. (2) near to unity.

In interpreting the data, it was necessary to calculate the short-circuit current for the *illuminated* section of specimen alone. With the experimental conditions used the path lengths of the carriers were small compared with the length of the illuminated section, so that any end effects due to carriers moving out of the illuminated zone may be ignored. The procedure used was to find the resistance r of a section of length equal to the focused image used for irradiation from resistivity measurements. Then from a measured voltage V across a load Z the short-circuit current is given by $i_s = V(Z+R)/Zr$ where R is the total specimen resistance. This expression of course allows for contact resistances (which are included in R) provided they are linear. As these resistances were kept low it was considered that any spurious effects arising at the contacts could be neglected.

Values of lifetime were also obtained from diffusion lengths measured by traversing a narrow slit of light over a fine tungsten probe and observing the photo-voltage generated at the probe as a function of the distance separating the light source and the probe. The theory of this measurement is described in the Appendix. The slit of light used was 15 μ wide, produced by reflecting optics so that no wavelength restrictions were introduced.* The radiation was chopped at 800 c/s and the photo-voltage measured with an amplifier tuned to this frequency.

---

* I am indebted to Mr. W. H. Mitchell of Telecommunications Research Establishment for cleaving and mounting the specimens. With care the contact resistances could be reduced to $\frac{1}{4}$ to $\frac{1}{2}\,\Omega$ for samples of only 0·1 mm^2 cross section.

† This amplifier system was designed by Mr. D. A. H. Brown of Telecommunications Research Establishment.

A third method of measuring lifetime was also used. This was direct observation of the build-up time of a photoconductive signal when a square pulse of light was allowed to fall on the specimen. The light source used was a spark gap, which after an initial delay of 0 2 μsec, was found by observation with a vacuum photocell to rise to half maximum value in 0·3 μsec (this time including any delays in the amplifiers used in the measurement), and to stay of fairly constant amplitude for about 4 μsec. The signal wave form was assumed to be exponential, and the lifetime taken to be the difference of the times required for the signal to reach $1-1/e$ of maximum for the PbS crystal and the vacuum photocell.

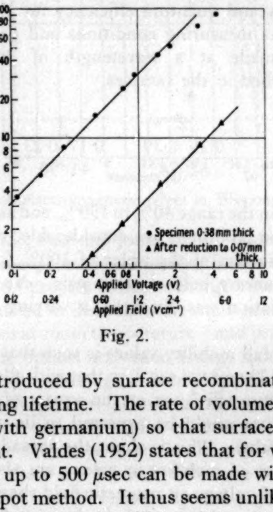


Fig. 2.

The complications introduced by surface recombination are ignored in all three methods of measuring lifetime. The rate of volume recombination in PbS is very high (compared with germanium) so that surface recombination should be relatively less important. Valdes (1952) states that for well-etched germanium measurements of lifetime up to 500 μsec can be made with reasonable accuracy using the travelling light spot method. It thus seems unlikely that recombination at cleavage surfaces of PbS crystals will produce appreciable errors in the measurements of such short lifetimes as are in fact observed.†

§ 4. EXPERIMENTAL RESULTS

The photo-current was found to vary linearly with applied voltage as shown in fig. 2, provided that the wattage dissipated in the specimen was not sufficient to cause appreciable heating. As shown in the lower curve of fig. 2 the range of linearity could be extended by reducing the specimen thickness (and so increasing its resistance) and at the same time mounting it on a suitable backing plate.

---

* This micro-illuminator was designed by Dr. D. G. Avery.

† According to Shockley (1950) the ratio (actual lifetime)/(volume lifetime) is $1-\tau s(2w^{-1}+2t^{-1})$. Taking $s=200$ cm sec^{-1} the highest value of τ in the present work (9 μsec) gives an error of 16%, with proportionately smaller errors for the shorter lifetimes.

4.1. *Quantum Efficiency*

In order to calculate the quantum efficiencies and lifetimes from eqns. (2) and (3), it is necessary to know the mobility and corresponding diffusion constant. There is still considerable uncertainty about mobility values for galena crystals. Unpublished Hall effect data by Dr. E. H. Putley of T.R.E. give an average mobility for either type of carrier of 400 $\text{cm}^2\,\text{v}^{-1}\,\text{sec}^{-1}$ for 6 samples of Joplin, 300 $\text{cm}^2\,\text{v}^{-1}\,\text{sec}^{-1}$ for 4 samples of Wisconsin and 540 $\text{cm}^2\,\text{v}^{-1}\,\text{sec}^{-1}$ for one sample from each of Saxony, Cornwall, Lead Hills, Urals, and Aberdeen, there being no significant difference between μ_e and μ_h. From these 15 samples the average mobility is 420 $\text{cm}^{-2}\,\text{v}^{-1}\,\text{sec}^{-1}$. Somewhat lower values have been given by Eisenmann (1940) but this is perhaps because some of the samples measured were polycrystalline. Thus from Hall effect data the average combined mobility is of the order of 800 $\text{cm}^2\,\text{v}^{-1}\,\text{sec}^{-1}$, and this value will be used to calculate both the quantum efficiencies and lifetimes.

Values of resistivity and quantum efficiency for Wisconsin galena are given in the table for various measuring conditions and parts of the specimen. All measurements were made at a wavelength of 2·0 μ, with approximately 10^{15} quanta sec^{-1} absorbed in the samples.

Specimen number	2	3	4	5
Resistivity (Ω cm)	0·29–0·39	0·17–0·23	0·23–0·45	0·50–0·76
Quantum efficiency (%)	52–74	41–90	69–150	40–48

These figures all lie in the range 40% to 150%, and as experimental inaccuracies and uncertainties in the mobilities are considerable, it seems probable that the quantum efficiency is in fact of the order of 100% for lead sulphide, i.e. each absorbed photon produces a hole–electron pair. No significant variation of η with intensity of irradiation was observed, *a* varying by less than 5% for 6:1 change of intensity.

The scatter in the Hall mobility values is such that 12 of Putley's 15 samples were within ± 40% of 420 $\text{cm}^2\,\text{v}^{-1}\,\text{sec}^{-1}$, so that with allowance for the dependence of the calculated quantum efficiency on the square of the mobility, 80% of the quantum efficiency values would be expected to lie within a factor of 2 up or down from the mean value. This is about the spread actually observed.

Relatively low values of mobility in galena are also indicated by the linear dependence of the PEM voltage on magnetic field shown in fig. 1. As a contrast the dependence for a germanium sample is shown. At high fields non-linearity is expected, due to a reduction in the PEM by a factor $1+\mu^2B^2$ (Aigrain and Bulliard 1953). This factor explains the germanium results fairly well if μ is taken as 0·45 $\text{m}^2\,\text{v}^{-1}\,\text{sec}^{-1}$, a value in fair agreement with that found by magneto-resistance measurements (see Moss 1952).

4.2. *Spectral Distribution of Sensitivity*

The spectral sensitivity curve was plotted for one sample using a Leiss double monochromator with LiF prisms. The results are shown in fig. 3. As will be seen from the lower curve the quantum efficiency is virtually constant from the near visible to 2·9 μ, the response falling rapidly at longer wavelengths, being down to half value at $\lambda_{1/2}=3{\cdot}02\ \mu$. There is thus no significant difference in the quantum efficiency at long wavelengths (in the tail absorption band) and at short wavelengths (in the main absorption band). Hence the 'tail' band cannot be regarded as an exciton band incapable of giving rise to photoconductivity in a single crystal, but must be regarded as arising from band to band transitions which produce free carriers. Furthermore, as the quantum efficiency in the 2 μ region is so high it is very unlikely that the photo-carriers can come from impurity centres, but they must come from the upper filled band of levels. From measurement of the optical constants of PbS of various impurity content Avery (1953) has also concluded that the 'tail' absorption band is intrinsic to the material. The $\lambda_{1/2}$ value found from fig. 3 coincides with the position of the absorption edge (Avery 1953, Paul and Jones 1953). Taking the optical activation energy as the quantum energy at this wavelength (see Moss 1952) its value is 0·41 ev. As the absorption process is not exciton formation (or even if it is, so little energy must be required to dissociate the excitons that they are dissociated at once by thermal energies at room temperature—and probably likewise at liquid air temperatures) then the optical activation energy cannot be significantly *less* than the forbidden gap width, although as a consequence of the Franck–Condon principle it may be slightly greater. However since the dielectric constant at radio frequencies* is approximately equal to that at optical frequencies, i.e. equal to the square of the refractive index in the non-dispersive region (Avery 1953), theory shows that there can be little difference between the thermal and optical activation energies (Moss 1952), and thus the gap width must approximate closely to 0·41 ev.

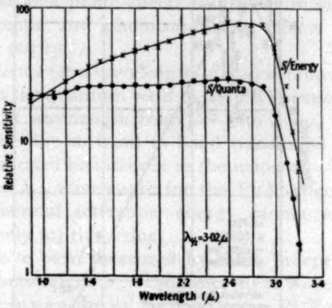


Fig. 3. Photoelectromagnetic effect in Wisconsin galena.

4.3. *Lifetime Measurements*

Using eqn. (2) lifetimes have been calculated for all the specimens of Wisconsin galena for which quantum efficiencies are given. In addition PEM measurements have been carried out on various specimens of Joplin and Broken Hill galena of considerably greater impurity content. It was not possible to measure photoconductivity in these latter samples, but on the basis of unit quantum efficiency (as established for the Wisconsin specimens) the lifetimes have been calculated from the absolute magnitudes of the PEM short-circuit currents using eqn. (1). The lifetimes for all samples are plotted against resistivity of the material in fig. 4, the difference points being for different parts of a given specimen. The values are seen to range from 6×10^{-10} sec to 9×10^{-6} sec, corresponding to diffusion lengths of about 1 μ to 100 μ. No variation of τ with wavelength was observed.

---

* See *International Critical Tables.*

Also shown in fig. 4 are lifetimes found from diffusion lengths which were measured by plotting the variation of photo-voltage at a metal probe as a slit of light was traversed over it. Some typical plots of photo-voltage against distance, which show that the theoretical exponential law is obeyed, are given in fig. 5.

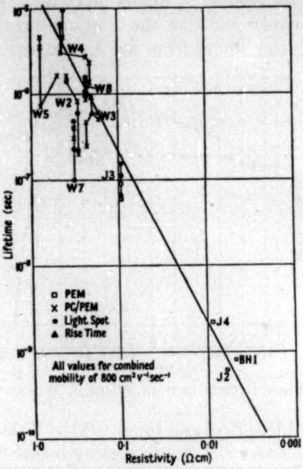


Fig. 4.

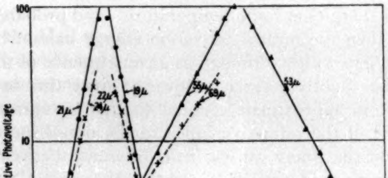


Fig. 5. Measurement of diffusion lengths by travelling light spot.

In fig. 4 there are also shown, for two specimens, values of response time obtained by direct observation of the rise of photo-current when 'square' pulses of light were allowed to fall on the specimens. It will be seen that the values of τ obtained by all three methods are in general agreement considering the scatter in values over different regions of the specimens, thus confirming the theories used for calculating τ.

Comparison with the straight line of slope 2 drawn on the graph shows that in general the data may be represented by the relation $\tau \propto \rho^2$, so that the lifetime varies inversely as the square of the concentration of free carriers. This indicates that the recombination of free carriers is a second order process, thus confirming Pincherle's (1951, unpublished) theoretical calculations that for PbS the Auger effect is the predominant mechanism. The straight line of fig. 4 represents a cross section for recombination of $2 \times 10^{-35} \times$ (carrier density) cm^2.

§ 5. CONCLUSIONS

It is concluded that bulk photo-effects take place in single crystals of PbS, both photoelectromagnetic and photoconductive effects being observable in specimens of adequate purity.

The quantum efficiency (at a wavelength of 2 μ) is found to be approximately unity in all cases. Also the quantum yield in the PEM experiment is shown to be virtually independent of wavelength from 0·9 μ to 2·9 μ, so that absorption in the 'tail' band must be due to band to band transitions yielding free carriers, and also the 'tail' band must be intrinsic to the material. The optical activation energy, found from the $\lambda_{1/2}$ wavelength for the PEM effect is 0·41 ev, and it is concluded that the thermal activation energy cannot exceed (and probably approximates very closely to) this value.

Carrier lifetimes have been measured by three independent methods, with reasonable agreement between them. Lifetimes up to 9 μsec have been observed, and it has been possible to measure values down to 6×10^{-10} sec by the PEM effect. For a wide range of samples the lifetime is roughly proportional to (resistivity)^2, indicating that the Auger effect is the predominant mechanism of recombination.

ACKNOWLEDGMENTS

Acknowledgment is due to the Chief Scientist, Ministry of Supply, and the Controller, Her Majesty's Stationery Office, for permission to publish this paper. I also wish to thank Dr. R. A. Smith, Dr. L. Pincherle and Mrs. A. M. Woodward for helpful discussions, and my other colleagues mentioned in the text who have assisted.

REFERENCES

AIGRAIN, P., and BULLIARD, H., 1953, *C.R. Acad. Sci., Paris*, **236,** 595, 672.
AVERY, D. G., 1953, *Proc. Phys. Soc.* B, **66,** 133.
BARDEEN, J., 1950, *Bell Syst. Tech. J.*, **29,** 469.
EISENMANN, L., 1940, *Ann. Phys., Lpz.*, **38,** 121.
GIBSON, A. F., 1952, *Proc. Phys. Soc.* B, **65,** 214.
MOSS, T. S., 1952, *Photoconductivity in the Elements* (London : Butterworths Scientific Publications).
MOSS, T. S., PINCHERLE, L., and WOODWARD, A. M., 1953, *Proc. Phys. Soc.* B, **66,** 743.
PAUL, W., and JONES, R. V., 1953, *Proc. Phys. Soc.* B, **66,** 194.
SHOCKLEY, W., 1950, *Electrons and Holes in Semiconductors* (New York : Van Nostrand).
VALDES, L. B., 1952, *Proc. Inst. Radio Engrs., N.Y.*, **40,** 1420.

APPENDIX

Theory of Travelling Light Spot Measurement of Diffusion Lengths

The dependence of floating potential and conductance of a probe on a semiconductor surface on the concentration of carriers in the neighbourhood of the probe has been discussed by Bardeen (1950), who finds $V_f = (\boldsymbol{k}T/\beta\boldsymbol{e}) \log (G/G_0)$

where V_f = floating potential, or open circuit voltage, G_0, G = normal and enhanced conductance respectively, $\beta\boldsymbol{e}$ = effective electronic charge as given by the rectification characteristic. Hence putting $G-G_0=\Delta G$,

$$V_f=\frac{\boldsymbol{k}T}{\beta\boldsymbol{e}}\log\left(1+\Delta G/G_0\right)=\frac{\boldsymbol{k}T\Delta G}{\beta\boldsymbol{e}\,G_0}\text{ if }\Delta G\ll G_0.$$

Thus for small increases of conductance the floating potential $V_f \propto \Delta G$. This condition is fulfilled fairly accurately for $V_f < 25$ mv for a typical value of $\beta=0{\cdot}5$. In the experimental measurements V_f was always less than 10 mv, so that the linear relation applies.

Bardeen (1950) has also shown that the increase in conductance ΔG is proportional to the added concentration of minority carriers near the probe, the result being sensibly unaltered when surface recombination of the normal order is considered.

Hence the open-circuit photo-voltage is proportional to the concentration of photo-carriers at the probe.

The continuity equations for the flow of the excess electrons, for example, are

$$\frac{\partial n}{\partial t}=-\frac{n}{\tau}+\frac{1}{\mathrm{e}}\nabla I=0\text{ at equilibrium.}$$

Now $I=\boldsymbol{e}\mu nF+\boldsymbol{e}D\nabla n=\boldsymbol{e}D\nabla n$ for zero applied field. For the one-dimensional case, i.e. ideally a thin filament, or in practice a point source of light near to a large area (p–n junction) collector,

$$I=+\boldsymbol{e}D\frac{dn}{dr}\text{ and }\nabla I=\frac{dI}{dr}=+\boldsymbol{e}D\frac{d^2n}{dr^2}.$$

For the two-dimensional case, i.e. ideally an infinitely thin sheet, or in practice a line source of light and a point collector,

$$\nabla I=+\frac{dI}{dr}+\frac{I}{r}.$$

For the three-dimensional case, i.e. point source and point collector,

$$\nabla I=+\frac{dI}{dr}+\frac{2I}{r}.$$

Hence in general for N dimensions,

$$\frac{d^2n}{dr^2}+\frac{N-1}{r}\frac{dn}{dr}-\frac{n}{L^2}=0\text{ where the diffusion length }L=(D\tau)^{1/2}.$$

For one dimension, $n\propto \mathrm{e}^{-r/L}$, and for three dimensions, $n\propto r^{-1}\mathrm{e}^{-r/L}$.

For two dimensions, the solution is a Bessel function, but an approximate solution is obtainable by putting $n=ur^{-1/2}$ where u is a new function. Then

$$\frac{d^2u}{dr^2}-u\left(\frac{1}{4r^2}+\frac{1}{L^2}\right)=0.$$

It is clear that for $r>L$ the $1/r^2$ term can be neglected, and we obtain $u\propto \mathrm{e}^{-r/L}$ and $n\propto r^{-1/2}\mathrm{e}^{-r/L}$. This is the expression used in analysing the data of fig. 5 where both log (photo-voltage) and log (photo-voltage $\times\, r^{1/2}$) are plotted, and where the distances for the signal to decay by e : 1 are marked on the graphs.

Extinction in X-Ray Diffraction Patterns of Powders

By A. R. LANG

Philips Laboratories, Inc., Irvington-on-Hudson, N.Y., U.S.A.

MS. received 25*th August* 1952, *and in amended form* 24*th August* 1953

Abstract. Errors are pointed out in analyses by Hall and Williamson and by Weiss of extinction in powders. Formulae are proposed giving a first-order correction for the reduction of reflecting power of small spherical crystals due to primary and secondary extinction. Neither formula accounts adequately for differences between observed and calculated intensities in Hall and Williamson's measurements on copper and aluminium.

---

§ 1. INTRODUCTION

It is nowadays possible to make precise measurements of the x-ray diffracting power of powders quite readily with the counter spectrometer. The observed diffracted intensity may, however, be influenced considerably by factors dependent upon the texture of the powder specimen. One such factor is extinction, and recent papers (Hall and Williamson 1951, Weiss 1952) have attempted to examine quantitatively its effect on the powder diffraction patterns of certain metals. In such studies it is necessary to consider how far the classical Darwin formulae are valid for the small particles which make up the powder specimen. It is also essential to ensure that these formulae are applied correctly. Hall and Williamson make errors in their analysis which largely invalidate the conclusions they draw. Weiss, too, makes errors, requiring his conclusions to be modified.

In both cases the errors principally concern the variation with diffraction angle of extinction losses. An attempt will here be made to state the correct formulae for this variation, applicable to the case when crystal-monochromatized radiation is used.

§ 2. PRIMARY EXTINCTION

Consider first primary extinction. If the value of the integrated reflection, in the absence of extinction, is denoted by ρ and the experimentally observed value by ρ', the well known Darwin correction for primary extinction is $\rho'/\rho=(\tanh mq)/mq$, where m is the number of planes occurring in regular succession and q is the amplitude reflected by a single plane, of spacing d,

$$q=\left(\frac{e^2}{mc^2}\right)NFd\lambda\times\text{polarization factor},$$

F is the structure amplitude of the plane and N is the number of unit cells per unit volume. It is convenient to denote the polarization factor by K where K is

1 or $|\cos 2\theta|$ and to introduce the term q_0 which is simply q without the polarization factor: $q_0 = (\mathbf{e}^2/\boldsymbol{mc}^2)\,N\,F\,d\,\lambda$. Then the integrated reflection from m perfect planes may be written $\rho'_K = MK \tanh Kmq_0$, in which M is a quantity independent of the polarization of the x-rays. For an incident beam monochromatized by reflection at a crystal with glancing angle α, averaging over the two mutually perpendicular directions of polarization gives

$$\rho'_{\mathrm{av}} = \frac{M}{1+\cos^2 2\alpha}\,[\tanh mq_0 + \cos^2 2\alpha\,|\cos 2\theta|\tanh(|\cos 2\theta|\,mq_0)] \quad \ldots(1)$$

in the arrangement, used in practice, when the normals to the monochromator and specimen reflecting planes are coplanar. In the case of very small mq_0 (i.e. no extinction coefficient) this expression reduces to

$$\rho_{\mathrm{av}} = \frac{M}{1+\cos^2 2\alpha}\,[mq_0(1+\cos^2 2\alpha\cos^2 2\theta)]. \qquad \ldots\ldots(2)$$

The above equations assume that the monochromator behaves as an imperfect crystal. If the monochromator were perfect then $\cos^2 2\alpha$ should be replaced by $|\cos 2\alpha|$ in all equations in which it occurs.* Expanding (1) and combining with (2) to obtain a first-order correction for primary extinction, we find

$$\rho'/\rho = 1 - \frac{m^2 q_0^{\,2}}{3}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right). \qquad \ldots\ldots(3)$$

This first-order correction will give values of the integrated reflection accurate to within about 1% when extinction losses do not exceed 10%.

§ 3. Secondary Extinction

Considering next secondary extinction; expressions for ρ' based on the Darwin model and expanded to a first order of approximation are, for symmetrical reflection at an effectively infinitely thick plate,

$$\rho'_K = \frac{Q_0 K^2}{2\mu}\left(1 - \frac{gQ_0K^2}{\mu}\right),$$

and for symmetrical transmission through a plate of thickness D

$$\rho'_K = Q_0K^2\frac{D}{\cos\theta}\left[\exp\left(\frac{-\mu D}{\cos\theta}\right)\right]\left[1 - gQ_0K^2\frac{D}{\cos\theta}\right],$$

in which

$$Q_0 = \left(\frac{\mathbf{e}^2}{\boldsymbol{mc}^2}\right)^2 N^2F^2\lambda^3/\sin 2\theta$$

and g is the secondary extinction coefficient. When the angular misorientations of mosaic blocks are distributed normally, with a standard deviation η in the plane of incidence, then $g = (2\pi^{1/2}\eta)^{-1}$. Averaging, as before, over the two directions of polarization we find, for the reflection case,

$$\rho' = \frac{Q_0}{2\mu}\left(\frac{1+\cos^2 2\alpha\cos^2 2\theta}{1+\cos^2 2\alpha}\right)\left[1 - \frac{gQ_0}{\mu}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right)\right];$$

---

* These polarization factors apply in all likely experimental arrangements, but not in the hypothetical condition of a point source of radiation and a crystal ideally focused on it. In this case a perfect, non-absorbing crystal will reflect totally both components of polarization, but an imperfect crystal will reflect these with intensities proportional to 1 and $\cos^2 2\alpha$, respectively.

hence
$$\frac{\rho'}{\rho}=1-\frac{gQ_0}{\mu}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right),$$
and for the transmission case
$$\rho'=\frac{Q_0D}{\cos\theta}\left[\exp\left(\frac{-\mu D}{\cos\theta}\right)\right]\left(\frac{1+\cos^2 2\alpha\cos^2 2\theta}{1+\cos^2 2\alpha}\right)\left[1-\frac{gQ_0D}{\cos\theta}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right)\right]$$
with
$$\frac{\rho'}{\rho}=1-\frac{gQ_0D}{\cos\theta}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right).$$

§ 4. APPLICATION OF EXTINCTION CORRECTIONS TO POWDERS

Now Hall and Williamson apparently omit the polarization factor from their primary-extinction correction. It is not clear what polarization factor Weiss uses in his primary-extinction correction formulae. However, more serious errors are to be found in these workers' secondary-extinction corrections. Hall and Williamson express their results in terms of the ratio f_{obs}/f, which is proportional to $(\rho'/\rho)^{1/2}$; f is Brindley's (1936) value of the atomic scattering factor for CuKα radiation of aluminium and copper, f_{obs} is Hall and Williamson's experimentally determined value, uncorrected for extinction and arbitrarily normalized with f at a certain reflection. The two secondary-extinction first-order correction formulae may be written
$$\frac{f_{\text{obs}}}{f}=1-\frac{gQ_0}{2\mu}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right),$$
and
$$\frac{f_{\text{obs}}}{f}=1-\frac{gQ_0D}{2\cos\theta}\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right). \quad\ldots\ldots(4)$$
But Hall and Williamson (1951, p. 943) put
$$f_{\text{obs}}/f=(1+kf^2)^{-1/2}\simeq 1-\tfrac{1}{2}kf^2,$$
thus absorbing all the angular factors into the constant k. Consequently the form of their curves showing the correction for secondary extinction (figs. 6 and 7, p. 944) is incorrect.

Weiss, on the other hand, takes the value of Q appropriate to a small powder volume instead of a small single-crystal volume (1952, eqns. (6) and (7), p. 554), and so his curve is also in error.

When applying the secondary-extinction correction to small spherical mosaics it is necessary to consider how formula (4) should be modified to allow for the spherical shape. Tables of the absorption factor for spherical crystals (Taylor 1944, Evans and Ekstein 1952) show that as long as absorption is quite small it is nearly independent of glancing angle and that the effective dimension $D/\cos\theta$ to be inserted in eqn. (4) is $0{\cdot}7D$, D being now the diameter of the mosaic. Thus
$$\frac{f_{\text{obs}}}{f}=1-0{\cdot}35gD\left(\frac{\boldsymbol{e}^2}{\boldsymbol{mc}^2}\right)^2N^2F^2\lambda^3\left(\frac{1}{\sin 2\theta}\right)\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right). \quad\ldots\ldots(5)$$
However, it must always be borne in mind that a small mosaic may not contain a number of individual coherent domains sufficient to give a smooth continuous distribution of orientations, and that the individual domains may not be large enough to possess an angular range of reflection small compared with the range of misorientations. In such circumstances formula (5) cannot be expected to apply.

Regarding primary extinction, it is rightly pointed out by Weiss that the primary-extinction correction formula to be employed should be derived for a small spherical perfect crystal rather than the thin plate of infinite lateral extent considered in the derivation of the Darwin formula. But the correction formula due to Ekstein (1951) which he quotes relates to peak reflecting power rather than to the integrated reflection. However, an indication of the ratio of the reduction of peak to that of integrated reflection can be obtained from comparison of the formulae for peak and integrated reflections given by Zachariasen (1945, pp. 131–134). In the case of primary extinction in reflection from a plate, the peak intensity P is proportional to $\tanh^2 A$, where A is equivalent to mq. Hence to a first approximation when the extinction is small

$$P = A^2(1-\tfrac{1}{3}A^2+\dots)^2 \simeq A^2(1-\tfrac{2}{3}A^2).$$

The integrated reflection ρ' is proportional to $\tanh A$, so $\rho' \simeq A(1-\tfrac{1}{3}A^2)$. Thus the relative reduction of peak intensity due to extinction is just double that of the integrated reflection.

In the case of transmission through a thin plate the peak intensity is proportional to $\sin^2 A$:

$$P = A^2(1-\tfrac{1}{6}A^2+\dots)^2 \simeq A^2(1-\tfrac{1}{3}A^2),$$

and the integrated reflection $= \int_0^{2A} J_0(x)\,dx$ is approximately proportional to $\int_0^{2A} (1-\tfrac{1}{4}x^2)\,dx$, which is proportional to $A(1-\tfrac{1}{3}A^2)$. So in this case the relative reduction is the same for both integrated and peak reflections. It may be suggested that the case of the sphere lies intermediate between those of reflection and transmission, and that for the integrated reflection the extinction correction is less than the Ekstein formula by a factor lying between one and two. However, the choice of this numerical factor is not critical in view of the inevitable variations of crystal shape and size present in the specimen. Retaining the Ekstein factor, the primary-extinction correction for a small sphere may be written

$$\frac{\rho'}{\rho} = 1 - \frac{7}{16}\left(\frac{e^2}{mc^2}\right)^2 N^2F^2\lambda^2D^2\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right),$$

and hence

$$\frac{f_{\text{obs}}}{f} = 1 - \frac{7}{32}\left(\frac{e^2}{mc^2}\right)^2 N^2F^2\lambda^2D^2\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right),$$

in which D is the diameter of the sphere. This may be compared with the Darwin formula (3), which, on substituting for mq_0 and expressing in terms of f_{obs}/f, becomes

$$\frac{f_{\text{obs}}}{f} = 1 - \frac{1}{6}\left(\frac{e^2}{mc^2}\right)^2 N^2F^2\lambda^2D^2\left(\frac{1}{\sin^2\theta}\right)\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right).$$

Here D is the thickness of the perfect region. It will be noted that the Ekstein formula does not contain the rapidly varying factor $\sin^2\theta$ in the denominator.

§ 5. SUMMARY AND CONCLUSIONS

To make clear the different angular factors, the above first-order correction formulae may be summarized as follows, setting $f=\tfrac{1}{4}F$:

(*a*) Darwin primary

$$\frac{f_{\text{obs}}}{f} = 1 - Af^2\left(\frac{1}{\sin^2\theta}\right)\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right).$$

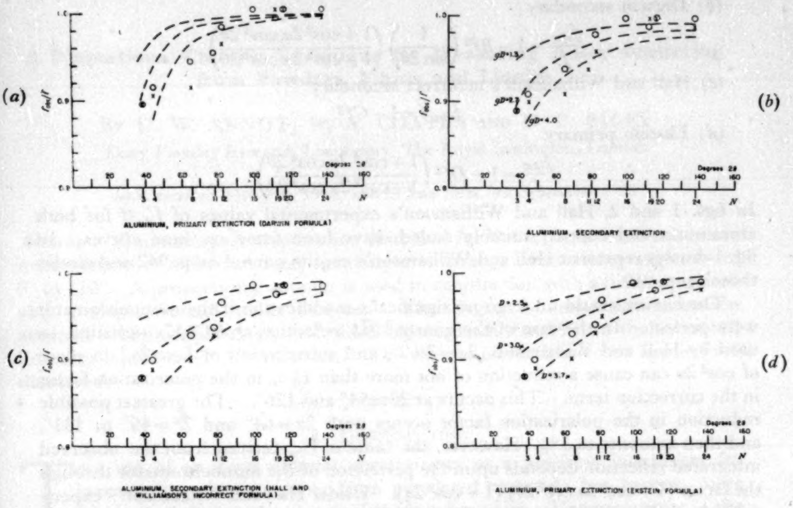


Fig. 1. Hall and Williamson's experimental *f* values for aluminium with extinction correction curves. (*a*) Darwin primary; (*b*) Darwin secondary; (*c*) Hall and Williamson's incorrect secondary; (*d*) Ekstein primary.

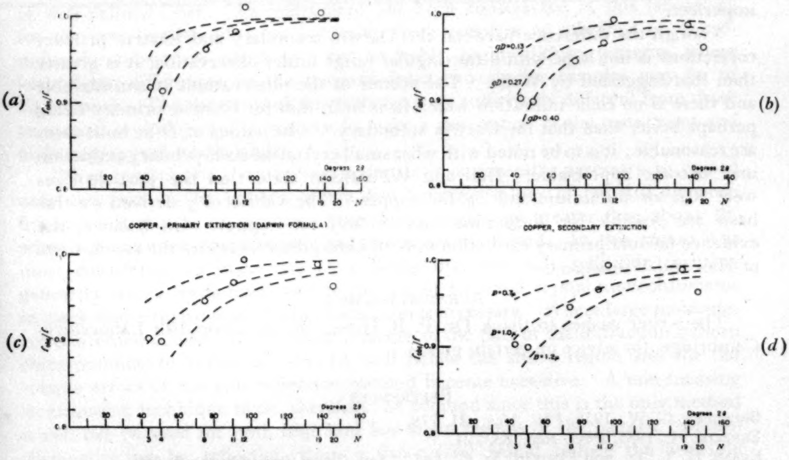


Fig. 2. Hall and Williamson's experimental *f* values for copper with extinction correction curves. (*a*) Darwin primary; (*b*) Darwin secondary; (*c*) Hall and Williamson's incorrect secondary; (*d*) Ekstein primary.

(*b*) Darwin secondary

$$\frac{f_{\text{obs}}}{f}=1-Bf^2\left(\frac{1}{\sin 2\theta}\right)\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right).$$

(*c*) Hall and Williamson's incorrect secondary

$$f_{\text{obs}}/f=1-Cf^2.$$

(*d*) Ekstein primary

$$\frac{f_{\text{obs}}}{f}=1-Df^2\left(\frac{1+\cos^2 2\alpha\cos^4 2\theta}{1+\cos^2 2\alpha\cos^2 2\theta}\right).$$

In figs. 1 and 2, Hall and Williamson's experimental values of f_{obs}/f for both aluminium and copper, suitably scaled, have been fitted to these curves. In fig. 1 crosses represent Hall and Williamson's results quoted on p. 942 and circles those on p. 949.

The curves would undergo no significant modification if the monochromator were perfect. In the case of the quartz $10\bar{1}1$ reflection and CuKα radiation, as used by Hall and Williamson, $2\alpha=26{\cdot}7°$, and substitution of $|\cos 2\alpha|$ in place of $\cos^2 2\alpha$ can cause a reduction of not more than $1\frac{1}{2}\%$ in the polarization factor in the correction term. This occurs at $2\theta\simeq54°$ and 126°. The greatest possible reduction in the polarization factor occurs with $2\alpha=64°$ and $2\theta=49°$ or 131°, and then amounts to 5%. However, the value of f_{obs} derived from the observed integrated reflection depends upon the perfection of the monochromator through the factor $(1+\cos^2 2\alpha\cos^2 2\theta)/(1+\cos^2 2\alpha)$. Under Hall and Williamson's experimental conditions substitution of $|\cos 2\alpha|$ in place of $\cos^2 2\alpha$ in this expression increases f_{obs} at the most by 2·6%, the greatest increase being in the region $2\theta=90°$; as the fit of f_{obs}/f with all the extinction-correction curves is thereby worsened it may reasonably be concluded that their monochromator was effectively imperfect.

Though the difference between the Darwin secondary and Ekstein primary corrections is not large within the angular range under observation, it is greater than that suggested by Weiss. The scatter of the observations is considerable and there is no clear indication which fit is best, that for Ekstein primary being perhaps better than that for Darwin secondary. The values of D in both cases are reasonable; it is to be noted with what small crystallite sizes primary extinction may become appreciable. Hall and Williamson state that the particle sizes were 50μ for aluminium and 2μ for copper. The values of g derived on this basis are possible for aluminium but unlikely for copper. On balance, the evidence favours primary extinction in both cases, thus reversing the conclusions of Hall and Williamson.

ACKNOWLEDGMENT

The writer wishes to thank Dr. P. B. Hirsch of the Cavendish Laboratory, Cambridge, for advice on certain points.

REFERENCES

BRINDLEY, G. W., 1936, *Phil. Mag.*, **21,** 778.
EKSTEIN, H., 1951, *Phys. Rev.*, **83,** 721.
EVANS, H. T., Jr., and EKSTEIN, M. G., 1952, *Acta Crystallogr., Camb.*, **5,** 540.
HALL, W. H., and WILLIAMSON, G. K., 1951, *Proc. Phys. Soc.* B, **64,** 937.
TAYLOR, A., 1944, *Phil. Mag.*, **35,** 215.
WEISS, R. J., 1952, *Proc. Phys. Soc.* B, **65,** 553.
ZACHARIASEN, W. H., 1945, *Theory of X-Ray Diffraction in Crystals* (New York: Wiley).

A Proportional-Counter Technique for Measuring X-Ray Scattering from Powders, Fibres and Liquids

By U. W. ARNDT, W. A. COATES and D. P. RILEY
Davy Faraday Research Laboratory, The Royal Institution, London

MS. received 24*th July* 1953, *and in final form* 3*rd September* 1953

Abstract. Counter spectrometers are described for the accurate semi-automatic measurement of diffracted x-ray intensity over a range of scattering angles from 0° to 130°. A proportional counter is used in conjunction with a linear amplifier and a single-channel pulse analyser. The degree of monochromatization attained with a copper target and nickel filter is such that 99% of the measured intensity is due to Cu Kα radiation. The various corrections to be applied are discussed.

---

§ 1. INTRODUCTION

THE use of an x-ray counter spectrometer as a precision instrument for investigating the scattering from powdered materials has become more widespread during the course of the last few years. However, most of the instruments described were designed primarily for recording Debye–Scherrer patterns from crystalline materials, where the main problem is that of the measurement of the positions, intensities and shapes of a relatively small number of well-defined lines. An instrument has been constructed in this laboratory for the investigation of x-ray scattering from powders or solutions of macromolecular substances which give rise to feeble, highly diffuse, patterns, where very slight differences in the patterns from two similar samples may be of significance. The demands of this technique are sufficiently different from those in normal crystallographic practice to make a description of the instrument and of its use appear worth while.

When investigating crystalline materials it is generally necessary to record only the part of the pattern which lies between values of $(\sin\theta)/\lambda$ between 0·2 and 0·8, that is, to spectra whose Bragg angles are greater than about 20° when the common characteristic radiations are employed. In this region it is most convenient to make use of a Bragg–Brentano side reflection focusing geometry which lends itself well to the use of a curved crystal monochromator in cases where rigorous monochromatization is necessary. Where large molecules are concerned considerable interest attaches to the part of the diffraction pattern corresponding to values of $(\sin\theta)/\lambda$ well below the above region and the flat sample errors of the side reflection method become excessive. A non-focusing transmission technique must, therefore, be adopted since this is the only method which can be used for both high and low angle regions of the pattern. Monochromatization by reflection from a flat crystal would reduce the available x-ray intensity too much and an alternative method, employing a proportional counter as a detecting device, has been developed to ensure that the recorded pattern is due to characteristic radiation alone.

§ 2. The Spectrometer

The instrument to be described was designed for recording scattering over the range $(\sin\theta)/\lambda = 0{\cdot}005$ to $0{\cdot}6$ using Cu Kα x-rays. The apparatus is shown schematically in fig. 1. The beam from the x-ray tube is collimated by the vertical slits S_1 and S_2. Two further slits S_3 and S_4 are mounted on the arm carrying the proportional counter C. S_1 to S_3 are all bilaterally adjustable slits, S_4 can be any one of a number of exchangeable fixed slits. The shaft on which the specimen P is mounted is geared to the counter arm so that the plane of the flat sample will always bisect the angle between the incident and the diffracted beam. The counter C is mounted with its axis vertical, i.e. parallel to the axis of the spectrometer. X-rays enter the counter through the bubble glass window W normal to the counter axis so that they pass through the region of maximum field uniformity in the counter. The monitoring Geiger counter G with its entrance slit S_5 is mounted on an arm pivoted beneath the aluminium foil F and is set so as to receive Kα x-rays diffracted by the foil. β-filters B_1 and B_2 are mounted in front of both counters.

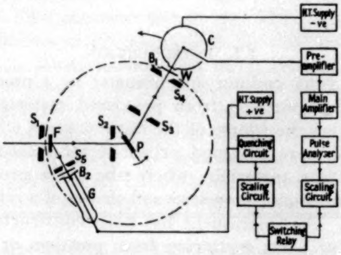


Fig. 1. Proportional-counter spectrometer (schematic).

§ 3. Monitoring

The monitoring arrangement is very similar to that described by Hall, Arndt and Smith (1949), where monitoring was carried out on a beam from another window of the x-ray tube. The present arrangement gives greatly improved stability, presumably because it is not affected by any slight movements of the focal spot which may occur. The intensity ratio between the two counters over ranges from 15 kv to 30 kv and from 5 mA to 20 mA drifts by less than 1% in 12 hours.

§ 4. The Specimen Holder

The specimen holder consists of a 1·5 mm thick stainless steel plate with two slots 12 mm wide; one of these holds the specimen under investigation, and the other is either left empty so that air scattering alone can be measured, or can contain a second specimen for comparison experiments. The holder is screwed to a vertical slide, operated from the control panel by a Bowden cable, so that one or the other of the two apertures can be brought into the x-ray beam. In order to load the holder it is clamped between two glass plates and the finely powdered material is filled into one of the two gaps and well rammed down with a spatula. Many materials have sufficient cohesion to remain as compacts when the glass is removed, but in some cases it is necessary to fit the holder with 'Cellophane' windows. Corrections must now be made for the scattering by the window in addition to allowing for air scatter. One piece of 'Cellophane' foil on each side is made to cover both apertures of the cell, thus ensuring that the orientation of both windows is the same with respect to the vertical axis, since thin foils always show some preferred orientation effects. When liquid samples are to be examined the slots in the specimen holder must be much narrower in order to avoid bowing inwards of the windows because of surface tension. Thin mica sheets (0·001 in.) are used when low-angle scattering from solutions is to be measured.

§ 5. THE TRAVERSE

The step-by-step traverse of the spectrometer is driven by a 100-step impulse motor. For greater ease in correcting the scattering curve by the various trigonometrical factors, it is convenient to make measurements at equal increments of $\sin\theta$ instead of θ, the geometrical factors being computed in advance and tabulated. The impulse motor is operated by a push button on the control panel and the number of steps which it has advanced is indicated on a Post Office message register connected in parallel with the motor. It has been found that when Fourier inversion of the corrected scattering curve is to be performed by means of the 3° Beevers–Lipson strips in order to derive a radial distribution curve (e.g. Riley and Arndt 1952), adequate representation of the details of the curve is given by making measurements at intervals of $(\sin\theta)/\lambda = 1/120$ over the range of the instrument. The number of steps of the impulse motor to reach these pre-determined positions is worked out once and for all.

§ 6. THE SLIT SYSTEM

Setting up of the instrument is very simple. With the aid of a 1 mm diameter steel rod accurately locating above the spectrometer axis, the spectrometer is moved and S_1 and S_2 are closed until the beam passes exactly over the axis of the spectrometer and its width is 1 mm. With the specimen slide containing a filled sample holder the counter is then set to a Bragg angle of about 10°. The widest of the fixed slits to be employed in scanning a particular pattern is inserted in front of the counter and S_3 is closed symmetrically until it just misses the diffracted beam. This can be done by noting changes in counting rate as S_3 is closed. There is a sudden rapid decrease when its jaws cut into the beam. The use of two slits on the counter arm improves the contrast, i.e. the ratio of counting rates with the specimen in and out of the beam, because the counter only 'sees' a very limited region of space in the neighbourhood of the specimen and thus does not record radiation scattered by air in other parts of the direct beam. In scanning a typical protein pattern it is our practice to use a counter slit S_4 of 1 mm width up to Bragg angles of 22°, the remainder of the scan where there is both less fine structure and less available intensity being carried out with a slit 2 mm wide. A sufficient overlap is allowed with the two scans to make accurate matching possible.

§ 7. MONOCHROMATIZATION

The principle of the method of monochromatization has already been described (Arndt and Riley 1952). The pulses from the proportional counter are fed into a high-gain linear amplifier. The amplified pulses are passed to a single-channel pulse amplifier whose channel is centred on the Cu Kα pulse amplitude. Figure 2 is a pulse height distribution curve obtained with the proportional counter when only crystal reflected Cu Kα radiation enters the counter. The distribution curve is approximately gaussian and the ratio of the root mean square deviation to the mean pulse height is 0·06. Apart from purely statistical fluctuations in the gas amplification factor (e.g. Wilkinson 1950, Lang 1952), the spread of the curve is due to three causes: (i) unequal gas amplification for ionization arising at different distances from the central wire within the counter, (ii) non-uniformity of the base line of the pulses owing to amplifier noise, (iii) drifts in the electronic circuits. The gas amplification of a proportional counter can be made independent of the position of the initial ionization by making the integrating and differentiating time constants of the amplifier equal, and at least as long as the collection time of the counter. However, since in x-ray diffraction applications fairly fast counting must be attempted, it is necessary to keep these time constants short to prevent build-up of pulses which would make their amplitude dependent on the counting rate. Time constants of only 1·0 μsec were therefore employed.

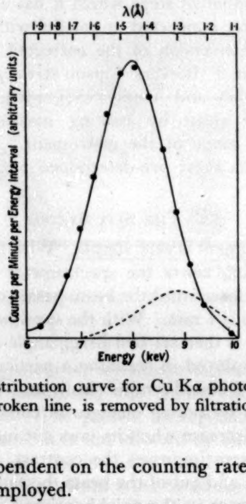


Fig. 2. Pulse amplitude distribution curve for Cu Kα photons. Cu Kβ peak, shown by broken line, is removed by filtration.

The signal-to-noise ratio in the amplifier could be improved by increasing the operating voltage of the counter, and so the gas amplification, thus permitting a lower amplifier gain. Once again, however, too great a gas amplification must be avoided in the interest of fast counting. When a proportional counter is used with a well collimated beam all the ionization is produced in a short section of the counter and the space charge thus generated reduces the gas amplification. This effect becomes more serious as the pulse size increases. A gas amplification of only about 100 was therefore employed, the external electronic amplification being 5×10^4.

The optimum conditions of operation of the amplifier and counter are fully discussed in the instruction manual of the pulse amplitude analyser type 1168 A. (Atomic Energy Research Establishment, Harwell, 1953.)

It will be seen that the resultant, approximately gaussian, distribution is considerably broader than the best that can be achieved with a proportional counter and this is due to the above compromises. Curran, Angus and Cockroft (1949) working with operating conditions and detecting circuits designed solely to obtain a maximum sharpness of the distribution were able to obtain a partial separation of the Kα and Kβ radiation of copper. Since in diffraction investigations stability must be maintained for prolonged periods, such methods could not be applied here.

The sharpness of discrimination against pulse heights departing from the mean, i.e. the width of the channel of the single-channel pulse analyser, is conditioned by stability considerations. The lower limit of the channel, the channel width, the gain of the amplifier and the polarizing voltage for the counter are all subject to small drifts and as a result the mean pulse height corresponding to Kα pulses is found to drift by about $\pm 1\frac{1}{2}\%$ in the course of a day, after an initial warming-up period. In order to reduce the counting rate variations within the channel to less than $\frac{1}{2}\%$ a channel width of ± 2 kev was chosen. This, of course, would not be sufficient to effect a complete separation of Cu Kα from other characteristic radiations which might be due to target impurities or fluorescence from the specimen. The Cu Kβ peak is shown by a broken line in fig. 2, but this is removed by filtration. The only other radiation which might be excited in the present work is the Lα radiation of tungsten arising from possible target contamination from the filament. The nickel absorption edge of the β-filter falls between the W Lα_1 and W Lα_2 wavelengths: the α_1-component which is about eight times as intense as the α_2-component is thus filtered out. Moreover, the target is kept clean so that this source of impurity should be of negligible importance. Previous work (Arndt and Riley 1952) has shown that with the arrangement described above the fraction of radiation detected which is not Cu Kα is about 1%.

§ 8. COUNTING LOSS CORRECTIONS

The losses of the proportional counter are due only to the finite paralysis time of the single-channel pulse analyser. With the sharp pulses employed this can be set to 5 μsec and the losses at any counting rate arising in practice can be neglected.

Hall and Williamson (1952) have considered the losses due to the finite resolving time τ_M of the monitoring counter. Under the above conditions their formula reduces to

$$N_L = -NM\tau_M/T \qquad \ldots\ldots(1)$$

where M is the number of counts of the monitoring before this stops a run of counts, T the time taken for this run and N the number of counts in this time of the measuring counter. If T is not allowed to vary by more than 3–4% by making occasional adjustments of X-ray tube current and voltage N_L can be regarded as a constant fraction of N and counting losses can be neglected altogether.

§ 9. ABSORPTION CORRECTION

The correction of the observed experimental scattering for absorption is easy to make when, as in the present case, the X-ray path through the sample is symmetrical.

Let I_1 be the observed scattering corresponding to a scattering angle 2θ when the specimen is in the beam and I_2 that when only the windows are in the beam. I_{sp} is the intensity scattered by the specimen, I_{c} the intensity scattered by the windows, I' the air scattering, μ_{sp} and μ_{c} the linear absorption factors of specimen and 'Cellophane' respectively, and t_{sp} and t_{c} the thickness of specimen and 'Cellophane' respectively.

Then $$I_1 = (I_{\mathrm{sp}} + I_{\mathrm{c}} + I') \exp\{-(\mu_{\mathrm{sp}} t_{\mathrm{sp}} + \mu_{\mathrm{c}} t_{\mathrm{c}}) \sec\theta\} \qquad \ldots\ldots(2)$$
and $$I_2 = (I_{\mathrm{c}} + I') \exp\{-(\mu_{\mathrm{c}} t_{\mathrm{c}}) \sec\theta\} \qquad \ldots\ldots(3)$$
so that $$I_{\mathrm{sp}} = I_1\, {}_{\mathrm{sp}}R_\theta - I_2\, {}_{\mathrm{c}}R_\theta \qquad \ldots\ldots(4)$$
where $${}_{\mathrm{sp}}R_\theta = \exp\{(\mu_{\mathrm{sp}} t_{\mathrm{sp}} + \mu_{\mathrm{c}} t_{\mathrm{c}}) \sec\theta\}$$
and $${}_{\mathrm{c}}R_\theta = \exp\{(\mu_{\mathrm{c}} t_{\mathrm{c}}) \sec\theta\}.$$

${}_{\mathrm{sp}}R_0$ and ${}_{\mathrm{c}}R_0$ are the ratios of incident to transmitted beams when $\theta = 0$, i.e. when the specimen is normal to the beam; they can be determined by absorption measurements. For this purpose a rock-salt crystal is mounted on the specimen spindle and the sample holder is transferred to a position in front of the counter. In this way the intensity of the direct beam is reduced sufficiently to allow the tube to be operated at normal power, and the beam can be monitored in the usual way.

Then $$\log R_\theta = \sec\theta \log R_0. \qquad \ldots\ldots(5)$$

It will be appreciated that the corrected intensity ordinates are rather sensitive to the value of ${}_{\mathrm{sp}}R_0$ so that considerable accuracy in the absorption measurements is necessary.

The value of I_{sp} obtained from eqn. (4) must, of course, be multiplied by a geometrical factor,
$$M = \cos\theta / (1 - \cos^2 2\theta) \qquad \ldots\ldots(6)$$
where $\cos\theta$ represents a correction for the change of volume as the specimen is rotated and $(1 - \cos^2 2\theta)^{-1}$ is the polarization factor.

Two special cases are of interest. The first is when no 'Cellophane' windows are required. In this case $I_{\mathrm{c}} = 0$ and ${}_{\mathrm{c}}R = 1$ when
$$I_{\mathrm{sp}} = I_1\, {}_{\mathrm{sp}}R_\theta - I_2 \qquad \ldots\ldots(7)$$
where I_2 now represents the air and slit scattering only. It has been found in practice that in this case the correction is so small and the monitoring maintains conditions to such a degree of constancy that for a given slit alignment I_2 can be determined once and for all at the sixty positions where measurements are taken and need only be checked at fortnightly intervals.

The validity of the corrections has been tested by examining samples of different thicknesses of the same material in specimen holders with and without windows. The resultant corrected curves were found to agree to about 1%.

The other case of importance is when dilute solutions of macromolecules are examined in which the interest is in the scattering at angles below 1° 2θ. In this case the 'blank' slot of the sample holder is filled with pure solvent and I_2 in eqn. (2) may be taken as including the scattering from the solvent. Further, in this range of angles R_θ may be taken as constant, and for moderately dilute solutions of proteins and similar materials the absorption of the solution and of the solvent alone is equal, so that
$$I_{\mathrm{sp}} \propto I_1 - I_2. \qquad \ldots\ldots(8)$$

§ 10. Low Angle Technique

The method has been extended for studies in the medium low angle range (scattering angles between 10^{-3} and 10^{-1} radian) in connection with studies on molecular shape factors in dilute solutions and on molecular packing in concentrated solutions. For very low angle work a subtractive technique is probably not applicable because of refraction (Yudowitch 1952) but in the above range values of radii of gyration of protein molecules have been obtained which agree well with published results (Ritland, Kaesberg and Beeman 1950, Anderegg, Beeman and Shulman 1952, private communication). A somewhat different type of counter traverse is used since at low angles it is not necessary to move the counter in a circular path around the specimen. A linear movement is sufficient and the specimen need not be rotated. The arrangement is shown schematically in fig. 3. The collimating system now consists of three bilaterally adjustable slits S_1, S_2, S_3, an aluminium foil F and a monitoring Geiger counter being employed as before. The specimen holder slide P is now attached to the front end of an evacuated brass tube T made from a length of waveguide. The removable backplate of this tube carries a 'Cellophane' vacuum window W′ inside which one of a number of backstops B can be fixed. The proportional counter C with its entrance slit S_4 is mounted on a carriage which can be moved in steps of 10^{-2} mm by a lead screw L, the latter being turned by an impulse motor M. No backstop is necessary for scattering angles up to 10^{-2} radian but at higher angles the 5 Å diffraction band from the 'Cellophane' of the rear window produces an unnecessarily high background counting rate unless the window is shielded from direct radiation. The slits, the evacuated tube and the counter traverse are mounted on an optical bench by means of saddles fitted with cross-slides. Waveguides of lengths 15 cm, 30 cm and 60 cm are provided for the investigation of the different angular ranges.

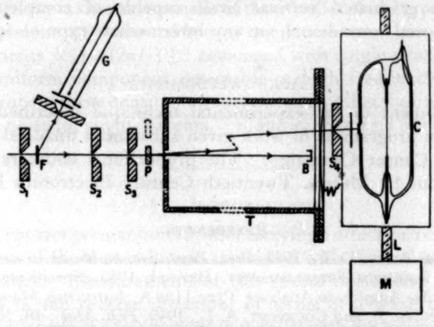


Fig. 3. Small-angle scattering instrument.

The particular types of correction required in low-angle scattering work have often been described elsewhere and will not be discussed here.

In this type of work a careful determination of the centre of the pattern is necessary and this is done by recording the intensities on both sides of the direct beam.

§ 11. Some Applications

The employment of a proportional counter technique for obtaining a monochromatic x-ray scattering pattern, coupled with effective monitoring of the incident beam intensity makes possible an accurate comparison of the scattering from two samples or the absolute determination of the scattering. The method has been found effective both for the recording of wide-angle patterns from non-crystalline samples and for studies at moderately low angles. Considerably greater speed and accuracy are achieved than are possible by any photographic technique employing monochromators. Thus it is possible in the course of a working day for one operator to derive two fully corrected scattering curves from powdered proteins or similar substances, covering a range of $(\sin\theta)/\lambda$ from 10^{-2} to 6×10^{-1}, the intensity ordinates being trustworthy to about 2%. The instrument can also be used for measuring the intensity diffracted by single fibres, or bundles of fibres. These are mounted in a holder consisting of a graduated vertical circle capable of complete rotation. In this way, equatorial, meridional, or any intermediate type of scattering can be investigated.

Acknowledgments

The development of the experimental technique described in this paper formed part of a programme of work given substantial financial support by the British Empire Cancer Campaign. The proportional counters were specially constructed for us by Messrs. Twentieth Century Electronics Ltd.

References

Arndt, U. W., and Riley, D. P., 1952, *Proc. Phys. Soc.* A, **65**, 74.

Atomic Energy Research Establishment, Harwell, 1953, Specifications and Manuals Section, Pulse Amplitude Analyser Type 1168 A, *Instruction Manual*, Issue 1.

Curran, S. C., Angus, J., and Cockroft, A. L., 1949, *Phil. Mag.*, **40**, 36.

Hall, W. H., Arndt, U. W., and Smith, R. A., 1949, *Proc. Phys. Soc.* A, **62**, 631.

Hall, W. H., and Williamson, G. K., 1952, *J. Sci. Instrum.*, **29**, 132.

Lang, A. R., 1952, *Proc. Phys. Soc.* A, **65**, 372.

Riley, D. P., and Arndt, U. W., 1952, *Nature, Lond.*, **169**, 138.

Ritland, H. N., Kaesberg, P., and Beeman, W. W., 1950, *J. Chem. Phys.*, **18**, 1237.

Wilkinson, D. H., 1950, *Ionization Chambers and Counters* (Cambridge: University Press), chap. 6.

Yudowitch, K. L., 1952, *Rev. Sci. Instrum.*, **23**, 83.

Measurements of Noise Spectra of a Point Contact Germanium Rectifier

By F. J. HYDE
Radio Research Station, Slough, Bucks.

Communicated by R. L. Smith-Rose; MS. received 8th July 1953,
and in final form 7th September 1953

Abstract. Measurements have been made of the excess noise produced by a point-contact germanium rectifier in the frequency range 0·117 c/s to 14 Mc/s at 29·5°c, with the reverse direct current I as parameter. The experimental noise spectral density is shown to consist of three well-defined types of component: (a) a basic component $\propto f^{-1}$ detectable over as many as seven decades of frequency, (b) two components $\propto [1+(2\pi f\tau)^2]^{-1}$ associated with single relaxation times $\tau=\tau_1$ and τ_2, (c) a uniform component detectable at high frequencies. There is no simple dependence of these densities on current, except that the uniform component is equal to shot noise for small currents. τ_1 and τ_2 are found to be dependent on I.

§ 1. Introduction

A POINT contact germanium rectifier carrying a direct current I exhibits noise in excess of the thermal noise associated with the real part of its internal admittance. Measurements of this excess noise have been made on a rectifier biased in the reverse direction, in terms of a noise current generator represented by a source of noise current in parallel with the internal a.c. impedance of the rectifier ($\rho\equiv\partial V/\partial I$ for the present range of frequencies). $\overline{\Delta i^2}$ is defined as the mean square noise current in a noise bandwidth Δf at frequency f, and experimental results are reduced to the form $\overline{i^2}=\overline{\Delta i^2}/\Delta f$, which is the mean square noise current in a 1 c/s bandwidth at frequency f. The equivalent noise e.m.f. has the mean square value $\overline{\Delta e^2}=\overline{\Delta i^2}\rho^2$ operating in series with ρ.

The present measurements refer to a particular commercial rectifier comprising a formed contact between a tungsten whisker and a piece of polycrystalline n-type germanium, which had been selected from many because of its stability as a noise source over long periods.

The noise of the collector contact of a type A germanium transistor is essentially that of a germanium point contact rectifier biased in the reverse direction. For such a collector contact there is experimental evidence (Ryder and Kircher 1949) that, in the frequency range 20 c/s to 15 kc/s, the mean square noise voltage (or current) varies as $f^{-1\cdot1}$. Similar measurements (Mooers 1949), in the frequency range 10 c/s to 16 kc/s, indicate that the exponent of frequency is not constant but varies from $-0\cdot5$ at 10 c/s to $-1\cdot4$ at 10 kc/s, passing through -1 at about 1 kc/s. Recent measurements (Montgomery 1952) on p–n junction type diodes and transistors between 1 kc/s and 200 kc/s show a dependence of noise density on frequency which is not significantly different from f^{-1} except, in some cases, in the region of 100 kc/s.

Other published noise measurements on germanium relate to current-carrying filaments (Herzog and van der Ziel 1951, Montgomery 1952). Herzog and van der Ziel made measurements on thin single-crystal filaments with large-area end contacts and found that the experimental excess-noise spectrum could be fitted by a formula

$$\overline{i^2} = \frac{AI^2}{f} + \frac{BI^2}{1+(f/f_0)^2} \qquad \ldots\ldots(1)$$

where $f_0 = 150$ kc/s and A and B are constants. Montgomery used filaments with potential electrodes and found a similar departure from an approximate f^{-1} dependence in the region of 100 kc/s.

For point contact germanium diodes Montgomery found that there was a general tendency for mean square noise current in a 1 c/s bandwidth at 1 kc/s to increase as I^2, but in limited regions the power of I could be considerably different from 2.

§ 2. EXPERIMENTAL METHOD

Noise measurements were made in the frequency range 0·117 c/s to 14 Mc/s. For frequencies up to 47·7 kc/s selective amplification of the noise was effected by amplifiers with switched components, to give predetermined control of resonance. Between 0·117 c/s and 184 c/s push–pull design was used because of the impracticability of decoupling at the very low frequencies, and the selectively was determined by a feedback circuit including unbalanced twin-T filters (Hyde 1952) ; from 11·5 c/s to 367 c/s a single-ended amplifier of the same type, and from 315 c/s to 47·7 kc/s a single-ended amplifier incorporating conventional inductor–capacitor tuning, was used. Beyond 60 kc/s a commercial communications receiver was used in conjunction with a tuned noise-input circuit. The Q-factors of all resonant responses were greater than 10.

The calibration source above 1 kc/s was a standard noise diode and below 1 kc/s a commercial sine-wave generator. Measurements in which the latter was used involved an accurate determination of the response of the amplifiers so that the noise bandwidths at each of the resonant frequencies could be determined.

Because of the sensitivity of the direct current–voltage relationship of the rectifier to temperature, all measurements were made with the rectifier in an enclosure, the temperature of which was maintained at 29·5 ± 0·1°c.

§ 3. EXPERIMENTAL OBSERVATIONS

The spectral density of the noise current $(\overline{i^2})^{1/2}$ is shown in fig. 1, with the direct current as parameter.

In fig. 2 the high-frequency spectral densities for $I = 25$ and 15 μA are shown. The lower levels (a′, b′) are those which obtained for the first few months during which readings were taken, and the higher ones (a, b) are the final ones after an abrupt change in noise had occurred, although the rectifier had in no way been overloaded. There was no measurable change in the magnitude of the low-frequency noise components nor in the direct current–voltage characteristic, concomitant with the change in high-frequency noise components.

The reverse d.c. characteristic is shown in fig. 3 (*a*) and the resistance determined by a.c. bridge measurement, as a function of I in fig. 3 (*b*). The a.c. resistance was measured to an accuracy within ± 1% as a function of frequency between 300 c/s and 600 kc/s for $I = 10$, 12·5 and 25 μA, and no dispersion was found.

§ 4. Analysis of the Spectra

The experimental curves can be closely synthesized, in the manner first suggested by Herzog and van der Ziel (1951), from the following semi-empirical expression for the noise current spectral density:

$$\overline{i^2}=C+A/f+B_1(1+\omega^2\tau_1^2)^{-1}+B_2(1+\omega^2\tau_2^2)^{-1} \qquad \ldots\ldots(2)$$

where $\omega=2\pi f$. C is the uniform component of noise current observed at high

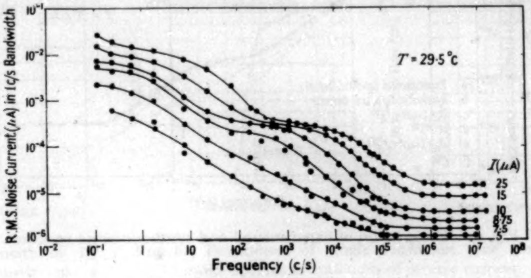


Fig. 1. Spectral density of noise.

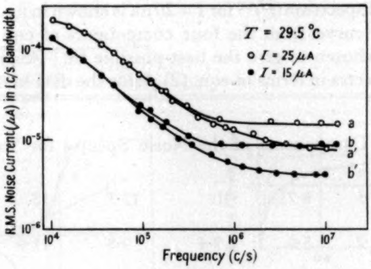


Fig. 2. High-frequency spectra showing intensity originally observed (a′, b′) and subsequent increase (a, b).

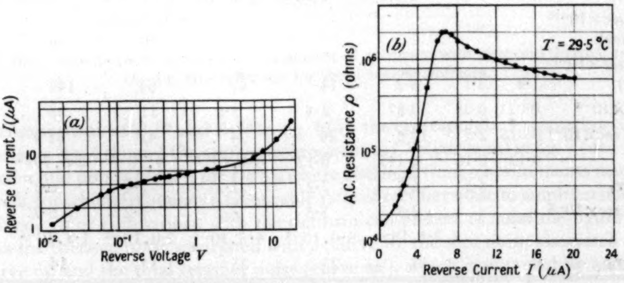


Fig. 3. (*a*) d.c. reverse characteristic ; (*b*) slope resistance as a function of reverse current.

frequencies. A/f is a basic component which contributes perceptibly to the noise over a wide range of frequencies, A being constant for a particular current. The two final components are associated with relaxation times τ_1 and τ_2 of the rectifier,

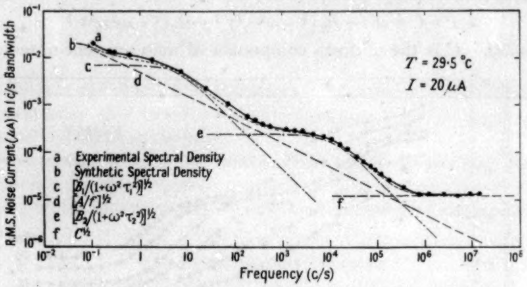


Fig. 4. Comparison of experimental and synthetic noise spectra.

B_1 and B_2 being constant for a particular current, and they are manifested in the spectra as 'bumps' above a smoothly decreasing curve.

The experimental spectrum $(\overline{i^2})^{1/2}$ for $I = 20\,\mu\text{A}$ is shown in fig. 4, and also shown is the synthesis of a curve from the four components of eqn. (2) with the six parameters suitably chosen to give the best possible fit. Analysis of the family of eight measured spectra in terms of eqn. (2) yields the data in the table.

Parameters of the Components of the Noise Spectra for Various Currents

Reverse current I (μA)	5	7·5	8·75	10	12·5	15	20	25
Reverse voltage V (v)	0·54	4·2	5·9	7·4	9·8	11·9	15	17·5
A.C. resistance (MΩ)	0·55	1·80	1·48	1·21	0·93	0·78	0·62	0·51
D.C. resistance (MΩ)	0·107	0·56	0·675	0·74	0·785	0·794	0·750	0·700
$C\,(\mu\text{A})^2\,\text{sec}\times 10^{-12}$ (original)						18	34	49
$C\,(\mu\text{A})^2\,\text{sec}\times 10^{-12}$ (final)	1·4	3·1	5·3	14	42	64	140	200
$A\,(\mu\text{A})^2\times 10^{-6}$	0·021	0·28	0·87	2·4	5·8	11	25	56
$B_1\,(\mu\text{A})^2\,\text{sec}\times 10^{-6}$	0·018	2·5	10·6	26	34	41	51	62
τ_1 (sec)	0·12	0·16	0·145	0·12	$6{\cdot}4.10^{-2}$	4.10^{-2}	$2{\cdot}6.10^{-2}$	$1{\cdot}3.10^{-2}$
$B_1/4\pi A\tau_1$		4·5	6·7	7·2	7·3	7·4	6·3	6·8
$B_2\,(\mu\text{A})^2\,\text{sec}\times 10^{-8}$			2·5	5·5	8·2	4·2	5·7	4·4
τ_2 (sec)			$2{\cdot}1.10^{-4}$	$8{\cdot}0.10^{-5}$	$4{\cdot}5.10^{-5}$	$2{\cdot}0.10^{-5}$	$1{\cdot}3.10^{-5}$	$8{\cdot}8.10^{-6}$
$B_2/4\pi A\tau_2$			11	23	25	15	14	7·4

The component involving B_2 is not detectable for $I < 8{\cdot}75\,\mu\text{A}$.

The various derived parameters are plotted in figs. 5 to 7 as functions of I. The $1/f$ *component* is shown in fig. 5 as $A^{1/2}$, the intensities of the *single relaxation*

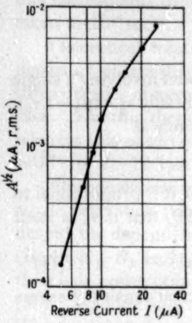


Fig. 5. Intensity of $1/f$ component as a function of reverse current.

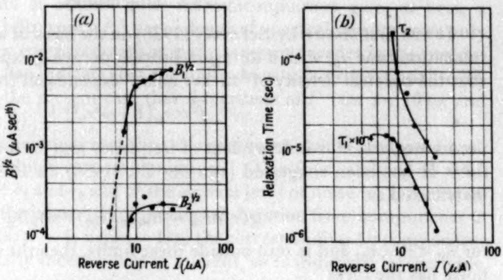


Fig. 6. Parameters of single relaxation time components of noise spectra as functions of reverse current: (*a*) intensity, (*b*) relaxation times.

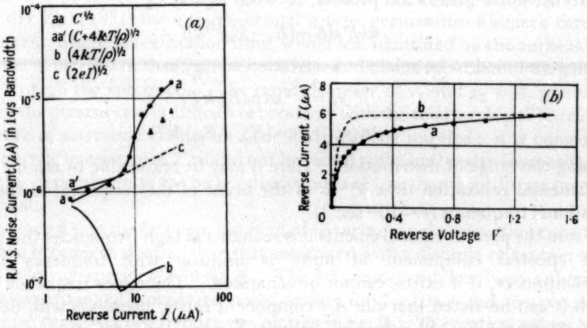


Fig. 7 (*a*). Comparison of uniform component of spectrum with shot and thermal noise, (*b*) d.c. characteristic for small reverse voltages.

time components as $B_1^{1/2}$ and $B_2^{1/2}$ in fig. 6 (*a*), the corresponding relaxation times τ_1 and τ_2 in fig. 6 (*b*) and the *uniform* component as $C^{1/2}$ (curve aa) in fig. 7 (*a*). The experimental points for $C^{1/2}$ for $I < 5$ μA were obtained from measurements made at 700 kc/s, for which frequency a favourable ratio of rectifier noise to amplifier noise is obtained. For comparison with the uniform component of noise the figure also shows the thermal noise associated with ρ (curve b), the shot noise associated with I (curve c), and the total rectifier noise (curve aa′), which is obtained by addition of measured excess noise and thermal noise on a power basis.

The parameters $B_1/4\pi A\tau_1$ and $B_2/4\pi A\tau_2$ given in the table are the ratios of the intensities of the two single relaxation time components to the A/f component at the frequencies ($1/2\pi\tau_1$ and $1/2\pi\tau_2$) at which the ratios attain their respective maxima.

§ 5. DISCUSSION

It was first shown by Schottky (1926) on the basis of a model involving a single relaxation time τ_0, with a diffusion process decaying exponentially as $\exp(-t/\tau_0)$, that the spectral density of 'flicker noise' depends on frequency as

$$\overline{i^2} \propto \tau_0[1+\omega^2\tau_0{}^2]^{-1} \qquad \ldots\ldots(3)$$

i.e. for $\omega\tau_0 \ll 1$, $\overline{i^2}$ is independent of frequency and for $\omega\tau_0 \gg 1$ it is proportional to f^{-2}. It has been suggested (van der Ziel 1950, du Pré 1950) that if τ_0 is itself distributed as

$$g(\tau_0)\,d\tau_0 = [\ln(\tau_{01}/\tau_{02})]^{-1} d\tau_0/\tau_0 \qquad \ldots\ldots(4)$$

for $\tau_{02} < \tau_0 < \tau_{01}$ and is zero outside these limits, then the spectral density of excess noise has the form

$$\overline{i^2} \propto (\tan^{-1}\omega\tau_{01} - \tan^{-1}\omega\tau_{02})/\omega. \qquad \ldots\ldots(5)$$

Then, for $\omega\tau_{01} \ll 1$, $\overline{i^2}$ is independent of frequency, for $\tau_{01}{}^{-1} < \omega < \tau_{02}{}^{-1}$ it is, to a close approximation, inversely proportional to frequency, and for $\omega\tau_{02} \gg 1$ it is proportional to f^{-2}. The particular distribution of τ_0 given in eqn. (4) is associated by these authors with a uniform distribution of activation energies, which govern the noise-generation process, between E_1 and E_2,

$$w(E)dE = (E_1 - E_2)^{-1} dE \qquad \ldots\ldots(6)$$

τ_0 being assumed to be related to activation energy as

$$\tau_0 = \nu^{-1}\exp(E_0/\boldsymbol{k}T) \qquad \ldots\ldots(7)$$

with

$$\ln(\tau_{01}/\tau_{02}) = (E_{01} - E_{02})/\boldsymbol{k}T \qquad \ldots\ldots(8)$$

defining the range of distribution. Here it may be reasonable to assume that the fundamental relaxation time ν^{-1} is of the order of the reciprocal of the atomic vibrational frequency ($\nu \sim 10^{12}\,\mathrm{sec}^{-1}$).

From the present measurements it is seen that at high frequencies the predominant spectral component of noise is uniform with frequency and the A/f component, if it exists, cannot be separated. The upper frequency limit, at which it can be stated that the A/f component exists, decreases with decreasing direct current. For all values of current, however, the A/f component can be specified as existent down to 0·117 c/s, the lowest frequency of measurement possible with the present equipment.

The A/f component of eqn. (2) is an idealized approximation to a component which will depart insignificantly from a dependence on f^{-1} over a wide frequency range if the defining relaxation times τ_{01} and τ_{02} of eqn. (5) are sufficiently disparate. For $\omega\tau_{01}$, $\omega\tau_{02}$ equal to unity the noise intensity will in both cases be 3 db down on that given by an exact f^{-1} dependence over the whole frequency range. This basic component of noise intensity contributes to the total noise in at least the frequency range $10^{-1} < f < 10^6$ c/s for a current of 25 μA, and departures from the ideal form are less than 1 db down at the extremes of frequency. The angular frequencies corresponding to a 1 db departure from an extensive f^{-1}

dependence at the frequency limits just stated are, on the basis of eqn. (5), equal to $3\omega_{01}$ and $\omega_{02}/3$ respectively, where $\omega_{01}=1/\tau_{01}$ and $\omega_{02}=1/\tau_{02}$. The range of relaxation times τ_0 is therefore at least $5{\cdot}3\times10^{-8}<\tau_0<4{\cdot}77$ sec and from eqn. (7) this corresponds to a range of activation energies of $0{\cdot}28<E_0<0{\cdot}76$ ev, ν being taken as 10^{12} sec^{-1}.

Theoretical treatments of contact noise in semiconductors give, at best, a departure of less than ± 1 db from a f^{-1} dependence of spectral density over only three decades of frequency compared with as many as seven decades in the present case. Existing theories also indicate that the noise density is proportional to I^2, whereas it is observed in the present case that A increases as $I^{7\cdot2}$ for $I<10\ \mu$A and as $I^{2\cdot8}$ for $I>12{\cdot}5\ \mu$A (fig. 5).

The behaviours of the parameters of the two noise components associated with single relaxation times exhibit certain similarities. The most striking is that, despite the dependence of τ_1 and τ_2 and of the general level of noise on I, the ratios $(B_1/4\pi A\tau_1,\ B_2/4\pi A\tau_2)$ of the intensities of the single relaxation time components at their half-power points $(2\pi f_1\tau_1=1,\ 2\pi f_2\tau_2=1)$ to the corresponding $1/f$ components $(A/f_1,\ A/f_2)$ are substantially independent of current, as is indicated in the table. This is regarded as good evidence that the same mechanism is responsible for both the $1/f$ and the single relaxation time components. For $I>10\ \mu$A, B_1 increases slowly with I, and B_2 is substantially constant, both τ_1 and τ_2 decreasing. For $I<10\ \mu$A, B_2 decreases rapidly and τ_2 increases rapidly, so that $B_2[1+\omega^2\tau_2^2]^{-1}$ is not detectable for $I<8{\cdot}75\ \mu$A; B_1 also decreases, but less rapidly than B_2, and τ_1 at the same time approaches an asymptotic limit of approximately 0·16 second. The general dependence of τ_1 and τ_2 on I is in contrast to the results of Herzog and van der Ziel (1951) for a single-crystal n-type germanium filament carrying direct current, where the relaxation time, which was identified by the authors with the lifetime of carriers in the filament, was found to be independent of the current flowing through the specimen. The rapid decrease of τ_1 and τ_2 with increasing I (or V) in the present case is difficult to reconcile with the relationship (7) because a dependence of activation energy on I (or V) is thereby implied: it is considered that the barrier temperature T would not be raised sufficiently with increasing I to explain this dependence as the power dissipations of less than $\frac{1}{2}$ milliwatt involved are too small.

It was found that the frequency spread of observed single relaxation time noise density was broader in all cases than that associated with a discrete relaxation time: this is in evidence specifically in fig. 4. The $1/f$ component has been subtracted from the experimentally observed noise density for $I=20\ \mu$A and the remaining excess-noise density at low frequencies plotted in fig. 8. Theoretical values based on various *relatively narrow* distributions of τ_1 of the form described by eqn. (4) are compared with it, all values being normalized at the half-power point for the noise density associated with the discrete relaxation time τ_1 (0·0266 sec). The distributions invoked are specified by values of $\log \tau_{11}/\tau_{12}$ of 0 (discrete relaxation time), 2, 3 and 4. The best fit to the experimental curve is given with $\log\tau_{11}/\tau_{12}=3$. The values 0, 2, 3 and 4 ascribed to $\log\tau_{11}/\tau_{12}$ correspond, on the basis of eqn. (7), to a narrow uniform distribution of activation energies E_1, $E_1\pm\boldsymbol{k}T$, $E_1\pm\frac{3}{2}\boldsymbol{k}T$ and $E_1\pm2\boldsymbol{k}T$, which are thermal spreads that might reasonably be assumed. This evidence is not inconsistent with the physical relationships implied by eqn. (7). A crucial test of this relationship would involve determination of the dependence of the experimental single relaxation times on temperature. It is considered that the interpretation of such experiments would be ambiguous, in that the single relaxation times τ_1 and τ_2 are dependent on I and V and it is not possible to maintain *both* I and V constant with variation of temperature.

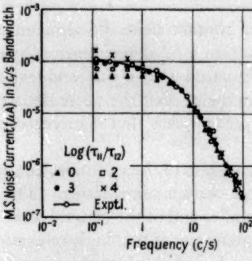


Fig. 8. Comparison of experimental ' single relaxation time ' noise density with values based on a discrete relaxation time τ_1 (0·0266 sec) and narrow distributions of relaxation time $g(\tau_1)\,d\tau_1=[\log(\tau_{11}/\tau_{12})]^{-1}\,d\tau_1/\tau_1$.

Beyond an upper frequency limit, which increases with increasing I, the spectral density of excess noise becomes uniform. For $I<7{\cdot}5\,\mu$A the rectifier d.c. characteristic (curve a) in fig. 7 (*b*) is seen to approach that of the ideal germanium diode (curve b); the synthetic curve is based on $I=6[1-\exp\{-eV/\boldsymbol{k}T\}]\,\mu$A, where $6\,\mu$A is the saturation current. In this region the excess noise is seen to be equal to the shot noise associated with the measured I. For $I>7{\cdot}5\,\mu$A the uniform component increases rapidly above shot noise. The originally measured values of uniform noise density for $\bar{I}=15$, 20 and $25\,\mu$A are also shown. That two distinct levels of uniform noise have been measured without any apparent change in the d.c. characteristic or in the frequency-dependent components of noise is evidence that the mechanism of uniform noise differs from that of frequency-dependent noise.

ACKNOWLEDGMENTS

The author wishes to thank Mr. R. E. Burgess of the Radio Research Station for his continued advice and encouragement and Messrs. J. G. Powell and G. E. Barnett for their assistance in construction of apparatus and in making measurements. The work described above was carried out as part of the programme of the Radio Research Board. The paper is published by permission of the Director of Radio Research of the Department of Scientific and Industrial Research.

REFERENCES

HERZOG, G. B., and VAN DER ZIEL, A., 1951, *Phys. Rev.*, **84**, 1249.
HYDE, F. J., 1952, *Wireless Engr.*, **29**, 85.
MONTGOMERY, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.
MOOERS, A. T., 1949, *Proc. National Electronics Conf.*, **5**, 17.
DU PRÉ, F. K., 1950, *Phys. Rev.*, **78**, 615.
RYDER, R. M., and KIRCHER, R. J., 1949, *Bell Syst. Tech. J.*, **28**, 33.
SCHOTKY, W., 1926, *Phys. Rev.*, **28**, 74.
VAN DER ZIEL, A., 1950, *Physica*, **16**, 359.

Recombination and Diffusion and Spread Echoes from the Ionosphere

By T. L. ECKERSLEY *

MS. received 10th June 1953

Abstract. This paper is concerned with the explanation of the spread F echoes often obtained on frequency-sweep records of equivalent height in the ionosphere, especially at night and in the auroral regions. It is shown in the Appendix that the production of a mountain range of echoes from a single incident pulse can be accounted for if there is a periodic structure of ionization in the F layer. There are two mechanisms of ionization in the ionosphere, namely the incidence of ultra-violet radiation from the sun, and the influx of a stream of charged particles that are deflected by the earth's magnetic field towards the geomagnetic poles. The ultra-violet light produces well defined layers, the E, F_1 and F_2 layers, with a single maximum of ionic density within them, and they may be satisfactorily explained by Chapman's theory, taking into account the recombination process but neglecting diffusion effects. In the case of particle ionization, however, the effect of diffusion becomes predominant. It is shown in the paper that the inclusion of the diffusion term in the differential equation of layer formation leads to the possibility of a periodic structure within the F layer of the kind needed to explain the spread echoes. This is because the differential equation has as its solution a doubly periodic elliptic function that provides the required periodicity of ionic density as a function of height. The form of the mathematics gives the essentially periodic nature of the solution without any detailed appeal to the initial conditions, but reference is made to the auroral curtains indicating the existence of sources of charged particles that maintain the periodic structure of ionization in between the curtains. Whereas the ultra-violet light causes layers to be formed that are many wavelengths thick, the particle ionization causes very thin layers only two or three wavelengths thick embedded in the background of ionization.

§ 1. INTRODUCTION

It is well known that the distribution of ionization with height in the ionosphere can be studied by measuring the time delay of an echo reflected from the ionosphere as a function of the radio frequency used. The time delay is expressed in terms of an equivalent height h', and the (h', f) curve obtained is often well defined and shows a rapid increase of height as the critical frequency is approached. For simplicity the complexities caused by multiple layers and multiple reflections and by magneto-ionic double refraction need not be considered here. Such a well-defined curve is an indication that the ionization then exists in the form of a relatively thick layer with a single maximum of density within it.

It is often found, however, that the echoes returned from the ionosphere are very spread, particularly at night-time and more especially in the auroral regions at any time of the day, giving rise to a diffuse and complex record in which it may be

---

* Private address : Weatheroak, Danbury, Essex.

difficult to distinguish any definite curve or critical frequency. Some excellent records of this kind taken at Spitzbergen have been published by Whatman (1949).

At first sight these spread echoes show some similarity to the long distance scattered echoes described by the author (1940), but whereas the latter are received at oblique incidence and need the use of high power, the spread echoes are reflected at practically vertical incidence and can easily be observed with a low power transmitter. Experimentally the scattered echoes are very difficult to balance on a pair of spaced frames when these are as much as 20 metres apart, whereas a balance can always be obtained on the spread echoes, showing that for them the bundle of rays reflected from the layer arrives practically vertically.

It is shown in the Appendix that such a system of spread echoes can be explained if within the layer there is a periodic structure of ionization, from which a single incident pulse is returned as a set of pulses that in practice will overlap and form a mountain range of the spread echo type. The presence of the spread echoes therefore suggests that the layer is no longer of the simple type with a single maximum of density within it, but that it has a multiple structure of the sandwich type with periodic variations with height in the layer.

The existence of these two types of layer can be explained in terms of two mechanisms of ionization, namely the incidence of ultra-violet radiation from the sun and the influx of a stream of charged particles that are deflected by the earth's magnetic field towards the geomagnetic poles. When the ultra-violet radiation alone is operative, the process of layer formation may be satisfactorily explained by including the effect of recombination but neglecting that of diffusion. The theory was worked out in detail by Chapman (1931), and the so-called Chapman layer is a relatively thick layer with a single well-defined maximum of density within it, such as the normal E, F_1 and F_2 layers.

When, however, particle ionization is also present, the Chapman theory must be supplemented, as the effect of diffusion can no longer be neglected; in fact the diffusion process plays a dominant role, and, as will be shown, it forms a system of thin layers only two or three wavelengths thick embedded in a background of ionization. The conditions under which a smooth layer of the Chapman type is formed are thus special rather than typical. When particle ionization is present, the ionosphere exhibits an instability that results through the diffusion process in the production of irregularities within the layer which take the form of a periodic structure of thin layers. Whereas the simple layer formed by the ultra-violet radiation from the sun will have a maximum density depending on the zenithal angle of the sun and so decreasing at night and in the polar latitudes, the irregularities produced by the particle ionization may occur at all times of the day and night and will be most common in the polar latitudes.

It is the purpose of the present paper to show how the inclusion of the diffusion term in the differential equation of layer formation leads to a steady state solution in which there exists a periodic structure, and hence to explain by the analysis in the Appendix the existence of the spread echoes. It will be shown that the solution of the differential equation is an elliptic function independently of the initial conditions, and it is one of the essential ideas of the paper that these initial conditions need not be known in detail. The principle is therefore established that even though the initial conditions are not given, the form of the mathematics implies that the solution must be periodic and that the spread echoes can thereby be explained.

The fundamental differential equation is set up in the next section, and in the succeeding sections the solution is given with the determination of the associated period. The particular case of a slab source is dealt with n some detail and the relation to the curtains observed in auroral regions is discussed.

§ 2. THE DIFFERENTIAL EQUATION OF LAYER PRODUCTION

The equation that forms the basis of the Chapman theory is

$$\frac{d\rho}{dt}=-\alpha\rho^2+q \qquad \ldots\ldots(1)$$

where ρ is the density of ionization at the time t, α is the recombination coefficient, and q is the rate of production of ions by the ultra-violet radiation.

When q also contains the ionization due to incoming charged particles and the effect of diffusion has to be included, eqn. (1) becomes

$$\frac{d\rho}{dt}=D\frac{d^2\rho}{dz^2}-\alpha\rho^2+q \qquad \ldots\ldots(2)$$

where z is height in the ionosphere and D is the diffusion coefficient given by

$$D=\tfrac{1}{3}vl=\tfrac{1}{3}l^2\nu_c \qquad \ldots\ldots(3)$$

where v is the velocity of the ions and l their mean free path, and ν_c is the collisional frequency.

The form of D may be modified by the presence of the earth's magnetic field, but this will not affect the form of eqn. (2) itself.

When the steady state with respect to time is established, $d\rho/dt=0$, and (2) becomes

$$D\frac{d^2\rho}{dz^2}-\alpha\rho^2+q=0. \qquad \ldots\ldots(4)$$

Before proceeding to the solution of this equation, we shall briefly discuss the assumptions upon which it is based. The distribution of density ρ in the ionosphere is mainly dependent on the distance from the centre of the earth. The electrons are considered to move in a flat space that is tangential to the sphere concentric with the earth at the point concerned. In this flat space the diffusion constant D is a function of the earth's magnetic field, since the movement of the electrons is modified by the presence of the field. This has been considered by Chapman and Cowling (1952) who discuss the effect of the earth's magnetic field on the diffusion of the electrons. In the differential equation, however, we can assume a definite constant value of D without needing to consider precisely how it is defined.

Some other aspects of diffusion are also left out of account. Thus a light gas will vary with the height in a different manner from a heavy gas, as the light particles tend to diffuse through the heavy ones. Thus the behaviour of the light particles depends upon the scale height H of the molecular particles in the atmosphere which is a function of their mass. In fact $H=\boldsymbol{RT/mg}$ and we have a distribution of particles of the form $\exp[-z/H]$ which would cause the lighter particles to be distributed in height in a different manner from the heavier particles.

The diffusion of the light particles through the heavy ones depends upon the difference between the scale heights of the two. If these scale heights are large enough, as they are in the practical case of the ionosphere, this factor is completely negligible, and we are justified in considering the differential equation as being of the form of eqn. (4), where ρ is one-dimensional depending only on z, and D is a simple constant.

Although there are possibly some details which have been neglected, this equation contains the central idea of the parts played by diffusion and recombination in determining the behaviour of the electrons in the ionosphere. This equation is essential for the understanding of the behaviour of the ionosphere, and its solution gives the major features to be considered.

§ 3. THE SOLUTION OF THE DIFFERENTIAL EQUATION

In the steady state with no ion production, eqn. (4) becomes

$$D\frac{d^2\rho}{dz^2} - \alpha\rho^2 = 0. \qquad \ldots\ldots(5)$$

In order to maintain a steady state in a region where there is no ion production, there must exist outside the region a place where q is not zero. This aspect of the problem will be considered later.

If we put

$$\alpha/D = S \qquad \ldots\ldots(6)$$

eqn. (5) may be written

$$\frac{d^2\rho}{dz^2} - S\rho^2 = 0, \quad \text{i.e.} \quad \frac{d}{dz}\left[\frac{1}{2}\left(\frac{d\rho}{dz}\right)^2 - \frac{S}{3}\rho^3\right] = 0$$

$$\text{i.e.} \quad \frac{1}{2}\left(\frac{d\rho}{dz}\right)^2 = \frac{S}{3}\rho^3 - \beta \qquad \ldots\ldots(7)$$

where β is an arbitrary constant.

The dimensions of the quantity S in eqn. (6) are L^3T^{-1}/L^2T^{-1}, i.e. L. Introducing a second arbitrary constant z_0, the solution of eqn. (7) is

$$z - z_0 = \int \frac{d\rho}{(\frac{2}{3}S\rho^3 - 2\beta)^{1/2}}. \qquad \ldots\ldots(8)$$

The integrand has three branch points given by

$$\rho^3 = 3\beta/S \qquad \ldots\ldots(9)$$

and no others, except possibly at $\rho = \infty$.

If we put $\rho = 1/\zeta$, the integral in eqn. (8) becomes

$$z - z_0 = -\int \frac{d\zeta}{\zeta^2(\frac{2}{3}S/\zeta^3 - 2\beta)^{1/2}} = -\int \frac{d\zeta}{\zeta^{1/2}(\frac{2}{3}S - 2\beta\zeta^3)^{1/2}} \qquad \ldots\ldots(10)$$

which shows that in terms of ζ there are four branch points, three of them lying on a circle in the complex ζ plane which are the roots of $\zeta^3 = S/3\beta$ and one at $\zeta = 0$ (or at $\rho = \infty$). These are representable on a four-sheeted Riemann surface in the complex ζ plane and are shown in fig. 1, where it will be seen that there are two independent closed circuits round the branch points, giving constants of integration I_1 and I_2, one of which is wholly real and the other wholly imaginary.

Thus it follows that

$$z - z_0 = -\int \frac{d\zeta}{\zeta^{1/2}(\frac{2}{3}S - 2\beta\zeta^3)^{1/2}} \pm mI_1 \pm nI_2 \qquad \ldots\ldots(11)$$

where m and n are integers.

If

$$-\int \frac{d\zeta}{\zeta^{1/2}(\frac{2}{3}S - 2\beta\zeta^3)^{1/2}} = f(\zeta)$$

then eqn. (11) may be written

$$f(\zeta) = z - z_0 \mp mI_1 \mp nI_2 \qquad \ldots\ldots(12)$$

and ζ is an elliptic function with periods I_1 and I_2 one of which is real and the other imaginary.

When $\beta=0$, the integral in (8) is very simple and there is no need to consider the elliptic function. It corresponds, however, only to the simple case of a single source, i.e. a single region where q is not zero and which maintains the ionic density in the region we are considering where q is zero. If we put $\beta=0$ in (8), it becomes

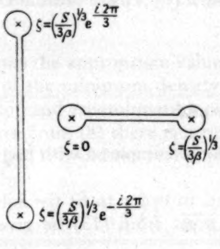
$$z-z_0=\int\frac{d\rho}{(\frac{2}{3}S)^{1/2}\rho^{3/2}}=-\frac{2}{(\frac{2}{3}S)^{1/2}\rho^{1/2}}$$

or

$$\rho=6/S(z-z_0)^2 \qquad \ldots\ldots(13)$$

which becomes infinite at $z=z_0$.

This obviously corresponds to a source at $z=z_0$ with the density decreasing monotonically at all points either greater or less than $z=z_0$ as shown diagrammatically in fig. 2. In the immediate vicinity of $z=z_0$, q is not zero and the above analysis does not apply. We may assume that in a small thickness $2d$ centred on $z=z_0$, q is not zero, but that outside this slab ρ is given by (13). We proceed therefore to consider how the solution in (13) can be joined on to the solution of the problem for such a slab source.



$\zeta=\left(\frac{S}{3\beta}\right)^{1/3}e^{\frac{i2\pi}{3}}$
$\zeta=0$
$\zeta=\left(\frac{S}{3\beta}\right)^{1/3}$
$\zeta=\left(\frac{S}{3\beta}\right)^{1/3}e^{-\frac{i2\pi}{3}}$

Fig. 1.

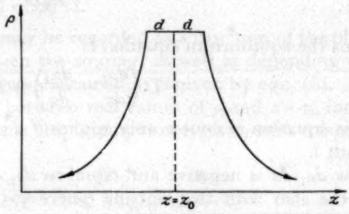


Fig. 2.

§ 4. The Slab Source

The source at $z=z_0$ where the infinity in ρ occurs according to the above analysis when $\beta=0$, may be regarded as a slab of small thickness $2d$ in which ionized particles are produced by ejections from the sun. Within this slab there must be no net gain of electrons with time, the increase due to the incoming particles being balanced by the loss of electrons by recombination and diffusion. In this region the differential equation with $q=0$ does not hold and we must make a direct study of the equilibrium of the slab to determine the relation between the intensity of supply and thickness of the slab.

We can do this because we know how the diffusion varies with the thickness of the slab, and we suppose that just outside the slab the mathematical relation in which the density varies as $1/(z-z_0)^2$ is true. The slab of finite thickness $2d$ is introduced to avoid the implication of an infinite source spread over an infinitesimal range of z if the differential equation were to hold right up to $z=z_0$.

The slab of thickness $2d$ is shown in fig. 3, and the solution of the differential equation is considered to hold to within a distance $d=z-z_0$ of the infinity at z_0

in this solution. Within the slab the equation does not hold because there is a supply of electrons q, and we can assume that the density ρ is nearly constant throughout the range $z=z_0 \pm d$. Considering an area A of the slab, the overall equilibrium of the layer can be determined as follows: The number of electrons lost per unit time from the slab is given by $-AD\,d\rho_1/dz + AD\,d\rho_2/dz$ where ρ_1 and ρ_2 are the densities at the two faces of the slab. The number of electrons lost by recombination is $A2d\alpha\rho^2$. The number of electrons in the slab supplied by the ionizing agency from outside per unit time is $Aq\,2d$.

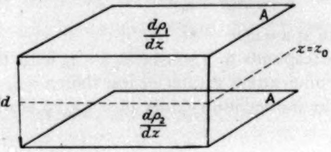


Fig. 3.

Thus the equilibrium equation is

$$-D\left(\frac{d\rho_1}{dz}-\frac{d\rho_2}{dz}\right)+2d\alpha\rho^2=q2d. \qquad \ldots\ldots(14)$$

This equation of course only applies if d is not large compared with the mean free path l.

Now $d\rho_1/dz$ is negative and equal to $d\rho_2/dz$, and in order to fit the solution inside the slab with that outside where $q=0$, ρ_2 must from (13) be given by $6/S(z-z_0)^2$ where $z-z_0=d$, so that

$$\left|\frac{d\rho_2}{dz}\right|=\frac{6}{S}\frac{2}{(z-z_0)^3} \qquad \text{at } z-z_0=d,$$

$$=12/Sd^3.$$

Thus (14) becomes

$$2D\frac{12}{Sd^3}+2d\alpha\frac{36}{S^2d^4}=2qd$$

i.e.

$$qd=\frac{12}{Sd^3}\left[D+\frac{3}{Sd}\right]$$

so that the total supply qd required to maintain a steady state is inversely proportional to d^3.

Physically it is fairly obvious that if the supply were very large, or infinite, a steady state could only be maintained if the electrons were also removed infinitely quickly. Thus the recombination $\alpha\rho^2$ and the diffusion $D\,d^2\rho/dz^2$ should also both tend to infinity. It is therefore possible to have a peak approaching an infinite value if the supply over a very thin stratum also tends to infinity. In other words there is no limit to the ionization gradient so long as there is no limit to qd, this gradient being limited purely by the supply.

§ 5. THE PERIODIC SOLUTION WHEN β IS NOT ZERO

Having seen that the special case when $\beta=0$ corresponds to the existence of a single source which maintains a monotonic decrease in density on either side of it, we now return to the general case when β is not zero. This must correspond to the case where there are a number of sources between each neighbouring pair of which the density distribution is given by a solution of (8) in which β is not zero and ρ is therefore a periodic function. It should be noted that as the differential equation is non-linear, we cannot build up the solution by the addition of the solutions for a number of single sources at $z=z_0$, $z=z_1$, $z=z_2$ etc. Rather we have to work out the completely general solution when β is not zero and apply it to each of the regions between the sources. In fact for each region there will be some appropriate value of β, and the solution between one pair of sources will not be analytically continuous with that for the regions beyond on either side.

We have then a picture of a series of slab sources the origin of which will be considered later. Between any neighbouring pair of sources the solution of the differential equation will be periodic, giving rise to a periodic structure of ionization between each pair of sources. In this structure the density does not drop to zero, but there are periodic minima, say ρ_0, associated with the value of β by the branch point condition in eqn. (9) which may be written

$$\beta = S\rho_0{}^3/3. \qquad \ldots\ldots(15)$$

Thus the appropriate value of β may be regarded as a function of the physical value of the minimum density between the sources, as well as depending on the diffusion and recombination coefficients contained in S given by eqn. (6).

Now from (8) there is a relation between real values of ρ and $z-z_0$ for $\rho>\rho_0$ given by (15), and the real period of the elliptic function is given with reference to fig. 1 by

$$I_1 = 2\int_{\rho_0}^{\infty} \frac{d\rho}{2(S\rho^3/3-2\beta)^{1/2}} = \sqrt{\frac{2}{\beta}}\int_{\rho_0}^{\infty} \frac{d\rho}{(\rho^3/\rho_0{}^3-1)^{1/2}}$$

using (15). Now put $\rho/\rho_0 = y$ so that $d\rho = \rho_0\,dy$. Then

$$I_1 = \rho_0\sqrt{\frac{2}{\beta}}\int_1^{\infty}\frac{dy}{(y^3-1)^{1/2}} = \rho_0\sqrt{\frac{6}{S\rho_0{}^3}}\int_1^{\infty}\frac{dy}{(y^3-1)^{1/2}}$$

i.e.
$$I_1 = \frac{\sqrt{6}}{S^{1/2}\rho_0{}^{1/2}}K \quad \text{where} \quad K = \int_1^{\infty}\frac{dy}{(y^3-1)^{1/2}}$$

which is a pure number. By putting $y^3 = 1+x$, this integral may be converted to a beta function that can be evaluated in terms of factorial functions giving $K = 2{\cdot}43$.

Millington (private communication) has derived this period by a somewhat different argument not explicitly invoking the knowledge that the solution is in the form of an elliptic function. His result agrees exactly with the above value, the period being given by the same integral. This forms a useful confirmation of the correctness of the analysis given. In particular it confirms that the period is inversely proportional to $S^{1/2}\rho_0{}^{1/2}$ and hence to $S^{1/3}\beta^{1/6}$ from eqn. (15).

The physical characteristics of the layer thus depend upon β, or alternatively the quantity β itself depends on the nature of the layer. It is clear that if we know the character of the F layer of the ionosphere we can follow up the analysis and see why the layer is often irregular in structure and how it may produce spread echoes. The physical analysis of the problem may help the understanding of it. On the other hand it is not essential, as the nature of the solution as a periodic structure is implicit in the mathematics. The form of the mathematical results does not depend upon a detailed knowledge of the initial conditions, i.e. on the physical conditions that determine the positions of the sources and the appropriate values of β in the regions between the sources. Whatever these initial conditions may be, the solution of the differential equation is periodic, and the generalized mathematical argument short-circuits the detailed physical analysis of the problem.

§ 6. THE SERIES OF SOURCES

It has been shown that the condition that β is not zero implies that there is a series of sources between which a periodic structure is maintained. Such a system of sources is known to be formed by particle ionization from the sun as is evidenced by the existence of auroral curtains. It is therefore reasonable to suppose that at times and places where spread echoes are observed, such a system of sources is present, implying that the ionizing agency is not only ultraviolet radiation, but that there is also a stream of incoming charged particles. Whereas, in their absence, the layer is of the normal Chapman type, their influence in the formation of a series of slab sources is to make diffusion a predominant factor in determining the structure of ionization in the regions between the slab sources.

§ 7. CONCLUSION

In § 4 it was shown that the initial conditions for a single source (when $\beta = 0$) imply a large production of ionization, defined by the rate of production of electrons q by an outside ionizing agency, over a short interval of height z. This corresponds to a single curtain of ionization. Experimental evidence given by Störmer (1938) of the production of auroral curtains shows that there can exist in the ionosphere sheets of ionization which lie along the direction of the earth's magnetic field. These sheets are formed by the influx of particles from the sun and their spacing is determined by quantum conditions governing the ejection of the particles from the sun into the earth's atmosphere.

In between two neighbouring sheets where the rate of production of ions given by q can be considered as zero, the distribution of ionic density ρ with height z is essentially determined by the solution of a differential equation of the type discussed above. Thus between the sheets there can exist a periodic structure of ionization of the kind required to explain the spread echoes by the argument given in the Appendix. The infinities which occur periodically in the elliptic function representing the solution of the differential equation can lie off the real z axis, and mathematically they then lie in the complex z plane. Thus along the real z axis between one auroral sheet and the next, the physical situation will be a series of maxima and minima forming a continuous periodically varying density–distance relationship.

Between each pair of auroral sheets there will be an associated elliptic function giving a periodic structure of ionization. Observations of the aurora show that the distances between successive sheets is not necessarily the same. This implies that the elliptic functions that represent the behaviour of the ionic density between the sheets are not related to one another and may have different periods depending upon the boundary conditions at the sheets. In other words, there may be a number of elliptic functions, each appropriate to a region between two neighbouring auroral sheets and describing the periodic structure of density therein, but not related to adjacent functions belonging to neighbouring regions.

Summarizing the theory that has been given above, it shows that when the rate of production of ionization from an outside agency is confined to a number of levels, the layer produced has a complex structure, and the mathematics definitely shows that the density distribution is given by a series of elliptic functions. The periodic nature of these functions implies that the layer is complex with a system of maxima and minima that accounts for the spread echoes as shown in the Appendix.

To find out the exact nature of the solution of the problem, the initial conditions would have to be known, and these depend upon the precise character of the ionizing agency, which is not known. But by examining experimentally the auroral effects, it would appear that the ionization is produced by incoming particles producing sheets of ionization about 1 km apart, as an examination of Störmer's photographs reveals.

The spread echoes are observed more especially in the reflections from the F layer of the ionosphere, and experimental evidence shows that the F layer irregularities are often produced during magnetic storms and in regions where the density due to ultra-violet light ionization is low. There is reason to believe, therefore, that the irregularities in the F layer are produced by particle ionization. The analysis given above bears out this contention and shows that the abnormalities in the F layer are of the sandwich type capable of producing the spread echoes. This structure may be controlled by the existence of auroral sheets of ionization, though in principle the period associated with the structure may be anything, depending upon the value of β in the integration process. In order to establish the existence of the periodic structure in the layer, it is only necessary to show, as has been done, that the density distribution is given as the solution of a differential equation which is of periodic type regardless of the initial conditions, which are unknown and can be left out of consideration.

ACKNOWLEDGMENT

My thanks are due to Mr. G. Millington for his help in the preparation of this paper.

REFERENCES

CHAPMAN, S., 1931, *Proc. Phys. Soc.*, **43,** 26.

CHAPMAN, S., and COWLING, T. G., 1952, *The Mathematical Theory of Non-Uniform Gases* (Cambridge : University Press), p. 332.

ECKERSLEY, T. L., 1940, *J. Instn. Elect. Engrs.*, **86,** 548.

STÖRMER, C., 1938, *Nature, Lond.*, **141,** 955; see also *Über die Probleme des Polarlichtes* from *Kosmischen Physik*, Vol. 1.

WHATMAN, A. B., 1949, *Proc. Phys. Soc.* B, **62,** 307.

APPENDIX

In order to show that a periodic structure within the layer will cause a single incident pulse to be reflected as a spread echo consisting of many pulses, we begin by considering the step function i_1 defined by

$$i_1 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp(2\pi i\nu t)\,d\nu}{\nu-\nu_0} \quad \ldots\ldots (\mathrm{A}\,1)$$

where the contour is chosen to include the pole at $\nu=\nu_0$ when it is completed by an infinite semicircle in the upper half of the complex ν plane as shown in fig. 4.

When $t>0$, $i_1=\exp(2\pi i\nu_0 t)$; but when $t<0$, $i_1=0$. This is illustrated in fig. 5.

We then build up a pulse of width τ by subtracting from this step function another one i_2 given by

$$i_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{A\exp\{2\pi i\nu(t-\tau)\}\,d\nu}{\nu-\nu_0} \qquad \ldots\ldots(\text{A}\,2)$$

which is zero when $t<\tau$ and equal to $A\exp\{2\pi i\nu_0(t-\tau)\}$ when $t>\tau$. This is shown with i_1 in fig. 6. i_1-i_2 is thus zero when $t<0$ and equal to $\exp(2\pi i\nu_0 t)$ when $0<t<\tau$, and will be zero when $t>\tau$ if A is chosen to be equal to $\exp(2\pi i\nu_0\tau)$, as shown in fig. 7. i_1-i_2 represents a pulse with infinitely sharp sides. We can obtain a pulse of any shape by building it up from elements of this kind as indicated in fig. 8. However, this simple picture is sufficient to express the main features of the pulse.

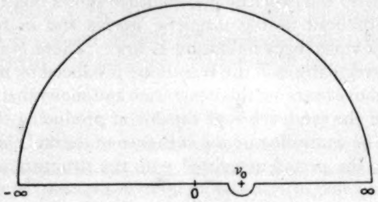


Fig. 4.

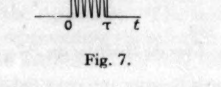


Fig. 5.

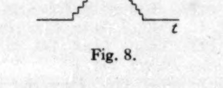


Fig. 6.

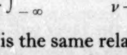


Fig. 7.

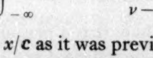


Fig. 8.

If the pulse travels to a distance x without dispersion it becomes

$$i_1-i_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\{2\pi i\nu(t-x/\mathbf{c})\}\,d\nu}{\nu-\nu_0} - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{A\exp\{2\pi i\nu(t-\tau-x/\mathbf{c})\}d\nu}{\nu-\nu_0}$$

and the pulse is the same relative to the time $t-x/\mathbf{c}$ as it was previously relative to the time t. It has therefore moved as a whole and is faithfully reproduced at a time $t=x/\mathbf{c}$.

If in a dispersive medium we have over a distance x a phase change ϕ, which must be a function of x and ν, we get for the form of the resultant pulse

$$i_1-i_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\{2\pi i\nu(t-x/\mathbf{c})-i\phi\}\,d\nu}{\nu-\nu_0} - \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\{2\pi i\nu(t-\tau-x/\mathbf{c})-i\phi\}\,d\nu}{\nu-\nu_0} \qquad \ldots\ldots(\text{A}\,3)$$

In fig. 9 the phase ϕ is shown as a function of the frequency ν together with the frequency spectrum of the pulse. The line AB represents a change in phase that is proportional to the frequency and corresponds to a wave travelling in free space. The delay of the pulse which the wave and side-waves produce is proportional to the slope of the line AB, i.e. to the rate of change $d\phi/d\nu$ of the phase with respect to the frequency.

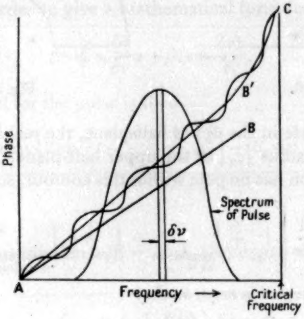


Fig. 9.

If the pulse is travelling in an ionosphere in which the density is uniform, the slope of the line increases and moves towards infinity as the critical frequency is approached, as shown by the curve AB′C. If the layer, like a sandwich, is made up of a number of subsidiary layers, the phase curve is no longer a smooth curve with an increase everywhere, but has a fluctuation on it as shown in fig. 9.

Analytically, for the case in which $d\phi/d\nu$ is constant, ϕ in (A 3) can be written as $\nu d\phi/d\nu$, the constant phase being omitted as it does not play any part in the integration. Then (A 3) becomes

$$i_1 - i_2 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\left\{2\pi i\nu\left(t - \frac{x}{\mathbf{c}} - \frac{1}{2\pi}\frac{d\phi}{d\nu}\right)\right\}d\nu}{\nu - \nu_0}$$

$$- \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\left\{2\pi i\nu\left(t - \tau - \frac{x}{\mathbf{c}} - \frac{1}{2\pi}\frac{d\phi}{d\nu}\right)\right\}d\nu}{\nu - \nu_0}$$

which represents a pulse at the time

$$t = \frac{x}{\mathbf{c}} + \frac{1}{2\pi}\frac{d\phi}{d\nu}.$$

This argument assumes that $d\phi/d\nu$ is constant over a sufficient range of ν for Kelvin's 'stationary phase' method of obtaining the group time to be applicable. The limits over which $d\phi/d\nu$ must remain constant in order that this method may be sufficiently accurate can be determined by assuming that the phase–frequency curve is linear up to a frequency ν_1 and then changes to another slope, as in fig. 10.

The main contribution to the integral must be in the neighbourhood of ν_0, and we wish to find, therefore, how far ν_1 must be removed from ν_0 in order that the contribution beyond ν_1 for a linear phase curve may be negligible. To do this, we consider a contour C_1 shown in fig. 11 which consists of the real axis from ν_1 to ∞, the infinite semicircle in the upper half-plane, the real axis from $-\infty$ to $-\nu_1$, and the semicircle of radius $|\nu_1|$ in the upper half-plane. The integrand of the type under consideration has no pole within this contour, so that

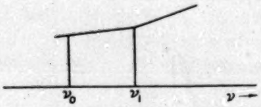


Fig. 10.

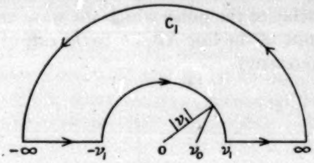


Fig. 11.

$$\frac{1}{2\pi}\int_{C_1} = \frac{1}{2\pi}\left[\int_{\nu_1}^{\infty} + \int_{\substack{\text{Semicircle}\\ \text{at } \infty}} + \int_{-\infty}^{-\nu_1} + \int_{\substack{\text{Semicircle}\\ \text{of radius } |\nu_1|}}\right] = 0.$$

Of these, the second integral is zero, so that

$$\frac{1}{2\pi}\left[\int_{-\infty}^{-\nu_1} + \int_{\nu_1}^{\infty}\right] = \frac{1}{2\pi}\int_{\substack{\text{Semicircle}\\ \text{of radius } |\nu_1|}} \qquad \text{......(A 4)}$$

where the semicircle is described in the conventional counterclockwise direction. On it we put $\nu = |\nu_1|\mathrm{e}^{i\theta}$, so that

$$\frac{1}{2\pi}\int_{\substack{\text{Semicircle}\\ \text{of radius } |\nu_1|}} = \frac{1}{2\pi}\int_0^{\pi} \frac{\exp\left\{2\pi i|\nu_1|\left(t-\frac{x}{\mathbf{c}}-\frac{1}{2\pi}\frac{d\phi}{d\nu}\right)(\cos\theta + i\sin\theta)\right\}d\nu}{\nu-\nu_0}.$$

Now
$$\frac{d\nu}{\nu-\nu_0} = \frac{|\nu_1|\,i\,\mathrm{e}^{i\theta}\,d\theta}{|\nu_1|\,\mathrm{e}^{i\theta}-\nu_0} = \frac{|\nu_1|\,i\,d\theta}{|\nu_1|-\nu_0\,\mathrm{e}^{-i\theta}}$$

so that
$$\left|\frac{d\nu}{\nu-\nu_0}\right| < \frac{|\nu_1|\,d\theta}{|\nu_1|-\nu_0}$$

and
$$\left|\frac{1}{2\pi}\int_{\substack{\text{Semicircle}\\ \text{of radius } |\nu_1|}}\right| < \frac{1}{2\pi}\int_0^{\pi}\frac{\exp\{-2\pi|\nu_1|\beta\sin\theta\}.|\nu_1|\,d\theta}{|\nu_1|-\nu_0}$$

where $\beta = t - \frac{x}{\mathbf{c}} - \frac{1}{2\pi}\frac{d\phi}{d\nu}$. Now this integral is less than

$$\frac{2|\nu_1|}{2\pi(|\nu_1|-\nu_0)}\int_0^{\pi/2}\exp\left\{-2\pi|\nu_1|\beta\frac{2}{\pi}\theta\right\}d\theta = \frac{1}{4\pi\beta(|\nu_1|-\nu_0)}[1-\exp\{-2\pi|\nu_1|\beta\}].$$

This must be small compared with unity, so that the condition is that $\beta(|\nu_1|-\nu_0)$ must be comparable with or large compared with unity.

Now the values of β which interest us and contribute to the impulse are those for which β is of the order of τ, as those for which $\beta \gg \tau$ do not contribute to the pulse. We can therefore write our condition as $1/\tau(|\nu_1|-\nu_0)$ small compared with unity.

The range of uniformity of phase as a function of ν is proportional to the inverse of the thickness of the pulse.

Fluctuations in the phase–frequency curve at a considerable distance from the actual frequency of the pulse may produce spurious echoes and distortion from a very narrow pulse, but will produce no appreciable distortion of a wide pulse.

If the layer density does not increase regularly, but in steps, we can have a phase function ϕ which also fluctuates about a mean value that increases linearly with frequency. In order to give a mathematical form for the function, we may take

$$\phi = \frac{\partial\phi}{\partial\nu}\nu + \beta_1 \sin\frac{2\pi\nu}{\nu_1}.$$

The dispersion integral for the pulse is then

$$i_1 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\left\{2\pi i\nu\left(t - \frac{x}{\mathbf{c}} - \frac{1}{2\pi}\frac{\partial\phi}{\partial\nu}\right) - i\beta_1 \sin\frac{2\pi\nu}{\nu_1}\right\} d\nu}{\nu - \nu_0}.$$

Now we can put

$$\exp\{-i\beta_1 \sin(2\pi\nu/\nu_1)\} = \sum_{-\infty}^{\infty} J_n(\beta_1)\exp(-i\,2\pi\, n\nu/\nu_1)$$

so that i_1 becomes

$$i_1 = \frac{1}{2\pi}\int_{-\infty}^{\infty} \frac{\exp\left\{2\pi i\nu\left(t - \frac{x}{\mathbf{c}} - \frac{1}{2\pi}\frac{\partial\phi}{\partial\nu}\right)\right\}\sum_{-\infty}^{\infty} J_n(\beta_1)\exp(-i\,2\pi\, n\nu/\nu_1)\, d\nu}{\nu - \nu_0}$$

$$= \frac{1}{2\pi}\sum_{-\infty}^{\infty} J_n(\beta_1)\int_{-\infty}^{\infty} \frac{\exp\left\{2\pi i\nu\left(t - \frac{x}{\mathbf{c}} - \frac{1}{2\pi}\frac{\partial\phi}{\partial\nu} - \frac{n}{\nu_1}\right)\right\} d\nu}{\nu - \nu_0}$$

with a similar integral for i_2 with t replaced by $t - \tau$.

The exponential in the integral is the same as before with $t - x/\mathbf{c} - (1/2\pi)\,d\phi/d\nu$ replaced by $t - x/\mathbf{c} - (1/2\pi)\,d\phi/d\nu - n/\nu_1$, where n is an integer. Thus if $t' = x/\mathbf{c} + (1/2\pi)\,d\phi/d\nu$ is the time at which the main pulse arrives, there will be extra pulses of magnitude $J_n(\beta_1)$ at the times $t_n = t' \pm n/\nu_1$, as indicated in fig. 12.

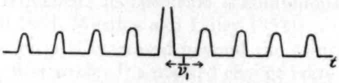


Fig. 12.

In theory, with an infinite series of fluctuations, there is an infinite series of pulses at the times t_n and these spurious pulses will be just as sharp as the original one. If there is a finite series of fluctuations, or if the fluctuations are slightly irregular, the pulses get rounded off, which is of course what happens in practice.

The separation between the pulses is $1/\nu_1$, and if this is small compared with τ, the thickness of the pulse, then the effect will be that of a distorted pulse and not that of a series of discrete spurious echoes. This is a check on the above conclusion that for a large distortion to occur the fluctuation must occur within a frequency ν_1 of the order of $1/\tau$.

A similar effect occurs if there is a fluctuation in amplitude as well as in frequency. This may well occur by partial reflections from thin layers such as may perhaps be considered to constitute the F layer. In this case the phase ϕ may be in the form

$$\phi = \frac{\partial \phi}{\partial \nu}\nu + i\beta_1 \sin\frac{2\pi\nu}{\nu_1}$$

so that the fluctuation of $i\phi$ is $-\beta_1 \sin 2\pi\nu/\nu_1$ and is real, and therefore represents an amplitude change. The analysis is exactly the same as before with $i\beta_1$ in the place of β_1, and the *n*th pulse is therefore proportional to $I_n(\beta)$ or $J_n(i\beta)$ and diverges to ∞ with *n*. It is therefore improbable that we can have an amplitude fluctuation over a very large range of frequency.

Complex values of β are also possible with a simultaneous amplitude and phase fluctuation if these are in phase. The mathematical formulation is more difficult if the phase and amplitude fluctuations are out of phase, but the splitting up into a mountain range of pulses is certain to occur. The main features of the solution are unaltered, and it seems highly probable that the spread of the observed pulses, which are all vertically propagated, is due to a phase and amplitude fluctuation, which in its turn is caused by a stratification of the F layer into regions where the density varies very rapidly in the vertical direction but only very slowly in the horizontal direction.

On the Mechanism of Choked Jet Noise*

By A. POWELL
Department of Aeronautical Engineering, University of Southampton

Communicated by E. J. Richards; MS. received 19*th May* 1953, *and in amended form* 7*th September* 1953

Abstract. The character of jet noise undergoes a marked change above choking, the noise due to turbulent mixing being dominated by a powerful whistle or screech whose wavelength is related to the regular shock wave spacing. The mechanism in two-dimensional flow is further examined (by the aid of a dynamic Schlieren apparatus), verifying the suggested mechanism and showing the similarity to that in axially symmetric flow where discontinuities in frequency, partly analogous to edge tones, occur. The resultant sound emitted as the periodic eddy system traverses the regular shock wave pattern is highly directional, producing a powerful beam at doubled frequency normal to the jet and an intense beam at eddy frequency in the upstream direction adjacent to the jet, resulting in fluctuations in jet velocity direction at the orifice which initiate new stream disturbances.

A gain criterion for the self-maintained cycle is given, enabling certain qualitative deductions concerning the intensity to be made, and use will be made of this in considering methods of reducing the noise level.

§ 1. INTRODUCTION—THE PHENOMENON

At a subsonic exit velocity the flow of an air jet, at a high Reynolds number and in the absence of pulsations or other periodic influences, is characterized by turbulent mixing commencing at the shear layers of the jet boundaries, and it is this which is responsible for the 'jet-noise' produced. This noise, best described as a 'hiss' for small jets or as a 'roar' for larger jets, has no discrete dominating frequency: its spectrum is continuous with a single, rather flat maximum (Powell 1951, Westley and Lilley 1952).

If the jet pressure ratio is increased beyond the critical (i.e. that at which a sonic exit velocity is first attained) a marked change takes place in the flow and also in the nature of the sound it produces. The flow is then 'choked' and expands on leaving the orifice, but after a certain distance (which is dependent upon the pressure ratio) contracts to its original diameter. Conditions are then very similar to those existing at the orifice itself, and the process then starts afresh and would repeat itself indefinitely if it were not for the turbulent mixing of the stream: even so several cycles are normally present, forming what has been called a 'cellular' pattern which is a prominent feature of Schlieren photographs of such a flow (see figs. 8 and 14 (Plate) for example), traces of as many as twelve having been observed (Powell 1951). Shock waves form during the contraction at the end of the 'cells', growing inwards in the upstream direction

---

* Formerly A.R.C. 15,623 F.M.1858, December 1952.

from the point of minimum area so as to form finally a conical structure. If the pressure ratio is still further increased a normal shock will form so as to remove the apex of the conical shock: most of the work of this paper refers to conditions prior to this latter development, which can be just distinguished in fig. 14.

The sound emanating from the jet undergoes a fundamental change. While the 'roaring' noise due to the turbulent mixing is still present, it may be almost completely dominated by a very powerful noise of completely different character. This might be described as a 'whistle' or 'screech', rather harsh and of a confused nature, becoming much more like a pure note, usually of increased intensity, over certain ranges of pressure ratio. This apparent stabilization occurs over several distinct regions, the pitch or frequency being different at each of these. These regions may be separated by a confused and unsteady whistling, but sometimes the note changes frequency quite abruptly. Measurements of the noise intensity showed a tendency for it to increase as a series of steps, and these are of special interest by virtue of their concidence with the regions of stabilization. This tendency may be seen in fig. 1 (reproduced from an earlier paper (Powell 1951)),

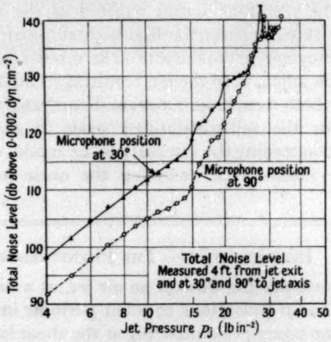


Fig. 1. Total noise level of 2 in. diameter jet exhausting to atmosphere.

which is for a jet of 2 in. exit diameter. (Owing to the response peculiar to the measuring instruments the points at which a near-tone dominated, i.e. the steps, may be slightly underestimated.) In other cases peaks, sometimes high, made an appearance at these steps. One such peak is evident towards the upper end of fig. 1. This figure also shows clearly how the noise level above choking (first occurring at a jet pressure p_j of about 13·2 lb in^{-2}, p_j being the total head of the jet relative to atmospheric pressure) increases very rapidly with jet pressure, depending upon it to a power of 6·0 (at 30° to the stream) or 9·0 (at 90°), roughly, in comparison with a mean value of about 3·6 below choking for the noise due to turbulent mixing. The difference in slopes of the noise energy graphs at the two positions indicates that the directional properties of the noise have also changed.

The above comments refer to flow from a circular orifice. The corresponding approximately two-dimensional case, that is the flow from a narrow rectangular orifice, has already received some attention (Powell 1951, 1953): here the powerful characteristic sound has a clearly defined frequency over the whole of the investigated ranges of jet pressure, the actual frequency depending upon the value of this pressure. It was suggested that the mechanism responsible for this characteristic sound was self-excited, and on making certain assumptions it was shown that the dimensions of various features of the flow and the highly directional sound field were compatible with that mechanism. It is the writer's aim in this present paper to examine the validity of those basic assumptions, and then to investigate the relation between the mechanisms in the two-dimensional and in the axially symmetric cases, the latter perhaps being important in view of the increasing pressure ratios of jet engines or rockets.

§2. EXPERIMENTAL APPARATUS

Mechanical Arrangements

The air supply for the jets was taken from a large reservoir, this being charged by a 40 h.p. electrically driven compressor. The supply entered the laboratory via a large main, the final run being of 3 in. diameter and containing the controlling gate valve. A Burgess type silencer 4 ft long, having a perforated inner wall and packed with sound-absorbent material between this and the outer pressure-tight case, was placed after the control valve to reduce as far as possible the effects of valve noise at the nozzle exit. The end of the silencer was fitted so as to take either a contraction to a 1 in. diameter nozzle, designed so as to avoid breakaway of the flow and producing a sensibly uniform velocity across the exit (Harrop 1951, Powell 1952 b), or to carry smaller nozzles having a rectangular exit.

Schlieren Apparatus

Photographs of the jet flows were taken by the Toepler–Schlieren method. A double-mirror arrangement, with 8 in. diameter 6 ft focal length mirrors, was used in conjunction with an auxiliary condensing lens adjacent to the light source. In this way a knife-edge window could be introduced to provide a sharp-edged effective source, and to remove unwanted images of the structure close to the actual light source. Adjustment of the image length at this point proved to be a convenient means of controlling the total illumination both when spark and when continuous light sources were used. A spark discharging between magnesium electrodes enclosed in a length of glass capillary tubing provided an ample light output with an effective duration of about a microsecond.

Dynamic Development

The problem of measuring the frequencies of certain disturbance motions in the jet flow arose, an upper frequency limit of at least 50 000 c/s being desirable. Since this is far above the working range of hot-wire techniques* and no suitable high speed camera was available (and the construction of one being a major task), it was decided to tackle the problem in a novel way. The method proved eminently suitable for its specific application, being particularly simple in construction and operation, and offers possibilities of development for other applications.

The existing Schlieren system was used, but instead of placing a photographic emulsion in the final image plane, the photo-cathode of a photomultiplier with a pinhole aperture was substituted. In this way the time fluctuation of a component of the mean density gradient in the flow could be obtained.

---

* *Note added in proof.* But see Kovásznay, 1953, N.A.C.A. Tech. Note 2839.

The photomultiplier used had a transparent photo-emissive cathode with a sensitivity of 28 μA per lumen. Eleven accelerating grids resulted in a current amplification of about $2{\cdot}2 \times 10^6$, the high voltage supply with a potential divider for these being as indicated in fig. 2, which shows the preliminary circuit arrangement. Since it was desired to maintain a sensibly constant response over the operating range of frequency, a cathode follower was used at the output. This was actually mounted within the light-tight container enclosing the photomultiplier, and with the circuit components indicated the estimated response was flat to within about 1 decibel at 150 000 c/s.

For this dynamic work a continuous light source was necessary, and a 380 w two filament 'solid source' gas-filled lamp was found quite satisfactory when run from a 24 v battery, only one filament being used.

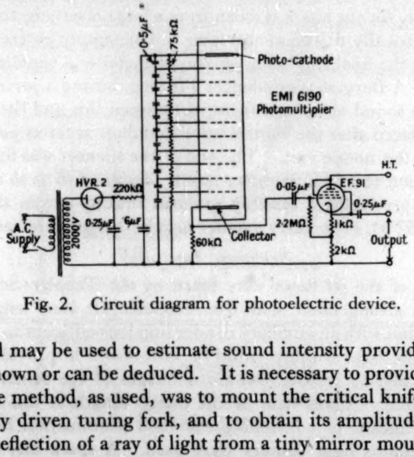


Fig. 2. Circuit diagram for photoelectric device.

The method may be used to estimate sound intensity provided the geometry of the field is known or can be deduced. It is necessary to provide some means of calibration: one method, as used, was to mount the critical knife-edge on the tip of an electrically driven tuning fork, and to obtain its amplitude of vibration by observing the deflection of a ray of light from a tiny mirror mounted on the fork, producing on a transparent scale a magnification, in this case, of 34. This method makes it unnecessary to depend upon calculations of the illumination, sensitivity of the photomultiplier or its aperture, or the electrical response so long as it is known to be flat. Some further comments and the method of calculating sound intensity are given in the Appendix.

Ancillary apparatus used in conjunction with the photoelectric device consisted of a wide range oscillator, a double-beam cathode-ray oscilloscope and a crystal microphone. Although the resonance of the latter occurred at about 17 000 c/s and the response fell above that, the sound levels were sufficiently high to make the use of an amplifier unnecessary, and the output was quite satisfactory for frequency comparisons, a principal consideration in the present analysis.

§ 3. FURTHER CONSIDERATION OF THE TWO-DIMENSIONAL CASE

Basic Assumptions and their Experimental Verification

The phenomenon occurring in the approximately two-dimensional case has been described in former papers, to which the reader is referred for details

(Powell 1951, especially 1953), in which a theory was put forward to explain the mechanism. Briefly, the observed regular stream disturbances were assumed to give rise to *stationary* sources of sound on traversing the cellular system with its shock waves, and several such sources would interact with each other to give rise to a highly directional sound field displaying certain *discrete* frequencies. On taking the frequency of the stream disturbances, which had a superficial resemblance to a Kàrmán Street, equal to the lowest radiated frequency of the sound, this directionality could be estimated for given cases and was in fact found to be in substantial agreement with that observed, i.e. a powerful lobe adjacent to the stream in the upstream direction, and a weaker one similarly downstream, whereas the first harmonic would have only a single major lobe normal to the stream. The powerful emission in the upstream direction was supposed to give rise to embryc disturbances at the orifice, which became amplified in the stream as they pass downstream. This work was based upon Schlieren photographs, and the frequency of the downstream radiation could not be ascertained with any confidence, due to its relative weakness.

Thus the theory depends upon the validity of the basic assumptions that (i) the disturbance is *exactly* equal to that of the sound passing in the upstream direction, (ii) the sources are stationary.

These assumptions were investigated in the following manner. The pinhole aperture of the photomultiplier (about 0·009 in. diameter) was arranged so that it was in the track of the disturbances on one side of the stream (the position being easily found from the existing Schlieren photographs or from direct observation since the envelope of the disturbances formed a 'ghost' on the aperture stop). The output from this was displayed together with that of the crystal microphone on the double-beam oscilloscope. It was arranged so that both beams were triggered by the disturbance wave form from the photoelectric device; this wave form was found to be quite regular and steady and, although harmonics were present, the fundamental was quite unmistakable. The intensity of this signal was clearly far above the 'noise' of the general turbulence.

When the microphone was located in the upstream direction close to the nozzle body its response, which was very nearly sinusoidal, was stationary on the oscilloscope screen and clearly of the same frequency as the disturbance fundamental. Any variation of frequency, even extremely small, would of course result in one wave form moving relative to the other. This equality existed over the whole range of pressure ratio investigated, corresponding to a frequency range of 18 000 to 42 000 c/s.

On placing the microphone in the downstream direction, where the theory indicates a frequency equal to the fundamental (Powell 1953), exactly the same result was obtained, the amplitude being smaller, as would be expected.

Finally, on placing the microphone in a position normal to the stream no significant response at the fundamental frequency was found, it being replaced by a strong signal at *precisely* twice that frequency.

Thus the over-riding assumption that the disturbance frequency is identical to that of the upstream source has been found true, and normal radiation found to be the first harmonic of this, exactly. Similarly the assumption of stationary sources has been verified, since if they were otherwise the Doppler shift of frequency would have been evident. This additional evidence adds a great deal of weight to the probability that the mechanism suggested is the correct one.

The Frequency of the Sound

The frequency of the sound has been determined from Schlieren photographs in the former papers. However, the photoelectric device makes much more convenient and accurate estimates possible: this could be done either by comparing a calibration signal from the oscillator with the disturbance wave form (having identically the same frequency as the fundamental of the sound) or by forming the first type of Lissajous figures. The latter method proved the more satisfactory, using the other as a check.

The variation of frequency with the jet pressure p_j (i.e. the total head of the jet less atmospheric pressure) found in this way is shown in fig. 3. An interesting feature was that the motion did not excite itself below a certain value of pressure (corresponding to a frequency of 42 000 c/s), either building up quite rapidly at that point or ceasing just as quickly, depending on the direction of the pressure change. This type of behaviour can be explained by the fact that the total gain round the 'circuit' of the self-excited mechanism is below unity on one side of this boundary and slightly above it on the other, so enabling an amplification until at a certain amplitude the gain becomes exactly unity and the motion is stabilized. These frequencies are up to about 10% greater than determined originally (Powell 1953) and a repeat determination (within a few hours) showed decreases of up to 4% on the later values. These discrepancies, considerably larger than experimental error, possibly arise through variations of humidity of the jet (no air driers were employed in either case), of temperature or of turbulence at the nozzle pipe entry, the layout up to this point being quite different in the earlier case. This reflects upon the susceptibility of the disturbance mechanism to conditions, perhaps not surprising since it is a self-excited process, and provides warning of the great care which must be exercised to maintain the good repeatability necessary in an extended research programme.

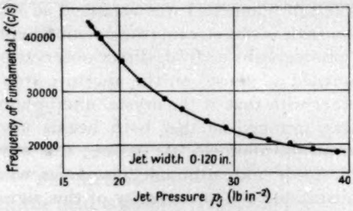


Fig. 3. Frequency of stream disturbance in two-dimensional case.

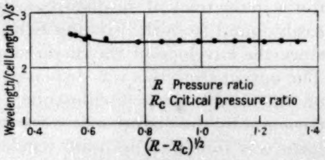


Fig. 4. Relation between fundamental radiated wavelength and cell length: two-dimensional case.

Now, if the disturbance motion becomes self-excited it will adjust itself so that the 'circuit' gain is a maximum (see Powell 1952 a for a general discussion on this aspect). The directionality of the radiated sound is acutely dependent on the cell spacing s, the disturbance spacing η and its speed M relative to sound, and is directly involved in this gain. It can be argued therefore that it will be near a maximum in the upstream direction. Since the intensity of the sound in the upstream direction depends upon $\cos\{n\pi(1+M)\,s/\eta\}$, where n is an integer (see Powell 1953), it follows that since $\eta = M\lambda$ (λ is the wavelength of the sound)

$$\lambda/s \rightarrow (1+M)/M.$$

Both M and s actually refer to the region at which the sound is generated, some distance from the orifice, but as an approximation s may be taken as the length of the first cell, which has been found to be $s/d = 1{\cdot}89(R-R_c)^{1/2}$, where d is the smaller dimension of the rectangular exit, R the jet pressure ratio, and R_c the critical value (Powell 1951). It is therefore of interest to see what values are actually found for λ/s, and these are shown in fig. 4. It is very nearly constant, $\lambda/s = 2{\cdot}5$, nearly, and this implies a disturbance Mach number of $\frac{2}{3}$ *. (Since s is somewhat overestimated, so is M.) One can therefore use the rough rule

$$\lambda = 5(R-R_c)^{1/2}\,\mathrm{d} \quad \text{or} \quad f = \frac{1}{5(R-R_c)^{1/2}}\left(\frac{c}{d}\right)$$

to estimate the order of wavelength or frequency of the fundamental sound, c being the velocity of sound.

The Intensity of the Sound

The sound intensity may be estimated from a knowledge of the geometry of the sound field and the associated density gradients; the details of the method used are given in the Appendix. The variation of intensity of the sound energy, obtained in this way, with jet pressure p_j was examined for a point in the upstream field, at about 30° to the axis and 2 in. from the effective source (i.e. between approximately 4 and 8 wavelengths distant), and is shown in fig. 5. Since the technique is quite novel and no convenient means of comparison are yet available, this determination should be regarded as being of a preliminary nature.

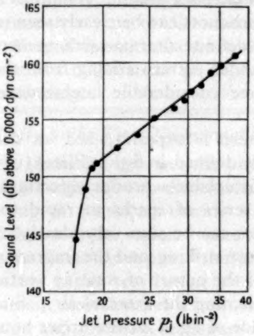


Fig. 5. Sound intensity in upstream field: two-dimensional case. Measurements by Schlieren and photoelectric apparatus.

The energy of the established motion follows the law $I \propto p^{3{\cdot}25}$, although there is as yet no theoretical estimate for comparison. At the lower limit the energy falls away very rapidly, the upper limit, if existing, being beyond the range investigated. The sound energy, expressed as a fraction of the kinetic energy of the jet, is considerably higher than that generated by the turbulent mixing, being of the order of 1%.

* Although conditions are widely different, this value happens to be the same as would result from applying Savic's (1941) result, applicable to neutrally stable disturbances in an incompressible laminar jet.

§4. THE PHENOMENA OF CIRCULAR JETS

It was the observation of stream disturbances and weak sound waves on Schlieren photographs of the flow of circular jets, i.e. axially symmetric, which led to the investigations of the flow under two-dimensional conditions. A theory meeting all the facts of that case has been put forward, and it will now be shown that the mechanism in the case of circular jets appears to be fundamentally of the same nature, and is responsible for the characteristic 'screeching' of such jets. As in the previous case, evidence is found from both Schlieren photography and from considerations of frequency. Except where clearly otherwise, the following observations refer to conditions at the steps, where, it will be recalled, the audible sound becomes more clear and steady (note that the photographs refer to a jet having an exit diameter of 1 in., whereas the noise measurements of fig. 1 refer to a 2 in. diameter jet: an exact correspondence between the two cannot thererore be expected).

*Results from Schlieren Photography**

Photographs of the flow have disclosed that two distinct phenomena may be present, on quite different scales: in addition to that form of disturbance in which we are primarily interested in the present discussion, there is also a phenomenon of much higher frequency and apparently different origin. The latter will be considered separately in a later section.

The large stream disturbances can be clearly seen in fig. 6 (Plate), where they can be seen to have a very definite alternate arrangement, and it will be noted that the later stages of the cellular pattern arising from the excess static pressure at the orifice have undergone considerable lateral deflections from their undisturbed central position.

This disturbance pattern is superimposed on the matrix of finer grained turbulence which is clearly defined in fig. 7 (Plate) (where the critical knife-edge had been reversed), the disturbances producing a zigzag effect. If the Schlieren screen illuminated by a series of sparks in rapid succession is observed, the passage of the disturbances can be seen very clearly. The lateral deflections of the cellular pattern are most striking, and the impression is often gained that the disturbances are in fact of the nature of rotating vortices, perhaps with filament of the turbulent flow entraining the external air. Such a case can be seen if a careful examination is made of fig. 7, at the upper boundary adjacent to the end of the third cell.

The general appearance of the stream is rather different when the critical knife-edge is parallel to the stream direction (fig. 8, Plate), the disturbance pattern then resembling the smoke filaments of sensitive jets as they form into vortices when the degree of instability is comparatively small (e.g. Brown 1935, fig. 5 (*f*)). A notable feature of the flow when these disturbances are present is the large angle of spread of the jet, 27° or 28°, in comparison with the normal angle of 16° or 18° as in fig. 9 (Plate), which illustrates a subsonic flow. It is interesting to reflect that this is the very property which gave rise to the so-called sensitivity of jets, i.e. the enhanced widening of the stream by the action of developing vortices. That the disturbances are of the nature of vortices has been suspected from the start, but in the majority of cases the development of the ragged turbulent boundaries towards that state from almost imperceptible undulations of the boundary has been insufficient to warrant the use of the word vortex. More and more evidence is being found associating the present phenomenon with the so-called sensitivity of jets, even to the point of the disturbance being sound induced, despite large differences in Mach and Reynolds numbers and the basic flows being turbulent in one case and laminar in the other.

---

* The features described in this section were first reported by Powell (1951).

The density gradients of the sound waves responsible for the 'screech' were insufficiently intense to be detected by the Schlieren apparatus in the foregoing cases, but may become so on increasing the pressure ratio to a higher value (p_j about 40 lb in^{-2}), although parts of the cellular and disturbance patterns then move from the field of view. The sound waves referred to are in phase opposition across the jet (fig. 10, Plate); one appears in the lower half of the photograph adjacent to the nozzle followed by another about one and a half cells from the exit, the one in the upper part of the picture lying midway between these positions. Photographs of a smaller jet have shown that these sound waves effectively emanate from the region at the end of the cellular pattern, where of course the disturbances are greatest (see Powell 1951).

In a few cases the disturbances and the radiated sound have been found to be not exactly out of phase, appearing more nearly in phase. It may be (for reasons of instability) that the disturbances are always approaching a state of being out of phase, the apparent deviations from this state being because the plane containing the disturbances is viewed obliquely, although no conclusive evidence has been found to show that the symmetric type of disturbance (i.e. involving toroidal vortices) does not appear. Even if the normal case is asymmetric, which appears to be so, symmetric disturbances might make an appearance over those regions where the motion is most unsteady.

Similarity to the Two-Dimensional Case

In view of the apparent similarity between the phenomena occurring in the flows from the circular orifice and in the two-dimensional case, the following tests were performed. Proceeding in a manner exactly similar to that used in the previous case, it was found that over the regions where the 'screech' was stable the wave forms associated with the disturbance procession and the sound rose far above the background 'noise' except at the very lowest pressure ratio and that (i) the sound radiated in the upstream direction has a frequency precisely equal to that of the disturbance procession and (ii) has a maximum in the upstream direction, (iii) is out of phase on each side of the nozzle for all the cases investigated and (iv) the sound radiated downstream has exactly the same frequency, (v) in the case $p_j \simeq 25$ lb in^{-2}, where the motion was particularly steady, the upstream sound was almost a pure sine wave falling away in intensity as a position normal to the jet is approached, at which position a relatively powerful 'beam' of sound at exactly double this frequency was found. In other cases the general effects were the same, but less easy to observe because of the presence of harmonics or some unsteadiness.

These features are identical with those found in the two-dimensional case: there is thus good evidence to suggest that the mechanism existing over the stable regions (i.e. over the steps discussed earlier) is the same as in that case. Since in the method of calculating the directionality of the sound field no assumptions were made restricting its applicability to the two-dimensional case, one would expect it to be valid in the present case. Taking the case at $p_j = 25$ lb in^{-2} it was found that the frequency was 5100 c/s, the disturbance wavelength 1·87 in. (obtained by the photoelectric device, using a pair of apertures (see Appendix), and checking with the value found from Schlieren photographs), and the cell spacing near the end of the cellular pattern 0·97 inch. If these values are substituted into the theoretical expressions for directionality of the fundamental frequency and the first harmonic (Powell 1953, eqns. (14)), it will be found that the fundamental has a maximum emission in the upstream direction, while the harmonic has a strong but narrow beam almost normal to the jet (fig. 11). This is in accord with the observations above for that case. At other conditions, of course, the directional properties may be somewhat different.

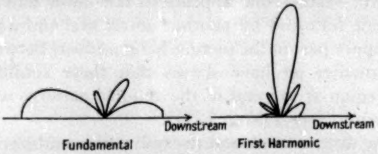


Fig. 11. Theoretical directionality for axially symmetric jet. $f = 5100$ c/s, $\eta = 1{\cdot}87$ in., $s = 0{\cdot}97$ in.

It is interesting to note that the stream disturbances were moving with a translational velocity of 0·711 times the speed of sound. There is some evidence to suggest that in some cases the disturbance speed may increase somewhat as they pass downstream.

The Frequency of the Sound

Whilst in the two-dimensional case the frequency and sound intensity vary steadily with the jet pressure, this is not so for the round jet. As noted in the Introduction, there are certain regions over which the motion appears stable and where the frequency and intensity vary steadily, but these may be separated by regions of instability, or the transition from one to another may be sudden. These features can be seen in fig. 12, showing the variation of frequency with jet pressure, where those regions have been labelled stages. Between stages 'a' and 'b' the motion is very confused although the amplitudes are still large. The transition from 'b' to 'c' is quite sudden, while a hysteresis is present at the next transition.

Although the frequency is discontinuous, its general trend is not unlike that of the two-dimensional case, and a plot of the ratio of the wavelength of the sound λ to the cell length s, $s/d = 1{\cdot}2(R - R_c)^{1/2}$, for the first cell (Powell 1951) is of interest (fig. 13), showing that the jumps of frequency do not result in a shift far from the values of the preceding stages (cf. the two-dimensional case, fig. 4). Note that s is the cell length adjacent to the exit, and at the higher pressure ratios it may be rather less than that at the other end of the cellular pattern, which is of the greater significance. Allowance for this effect, a decrease in s up to 20% (see for example Powell 1951, fig. 7) would result in the ratio λ/s being more nearly constant, again at about 2·5. Rough rules for the wavelength of the sound, and its frequency then are

$$\lambda = 3(R - R_c)^{1/2} d \quad \text{and} \quad f = \frac{1}{3(R - R_c)^{1/2}}\left(\frac{c}{d}\right),$$

in the previous notation, with d the jet diameter, independently of the stage. The latter is an important observation.

On the Discontinuities of Frequency

The reasons for the existence of the discontinuities of frequency can be explained qualitatively by reference to the conditions governing the self-maintained process. The similarity of the mechanism to the edge tone phenomenon has already been pointed out (Powell 1952 a), the criteria being of similar form. The first concerns certain phase relationships. Let the distance between the effective source of the sound and the orifice be h. Then the time taken for a disturbance to reach that source from the orifice where it was created by a passing sound wave (according to the suggested theory for the two-dimensional case, Powell 1953, which has been found to be present in this case also) and for a sound wave of phase similar to that just generated to reach the orifice will be

$$\int \frac{dh}{Mc} + \frac{h - l\lambda}{c},$$

where M is the Mach number of the disturbances relative to the *ambient* speed of sound c. The term $l\lambda$ is introduced to take into account the fact that if l wavelengths exist between the orifice and the source, then the phase of sound wave in which we are interested will reach the orifice l cycles before the one identified specifically with the disturbance. In the meantime θ cycles of the stream disturbance will have been created, so that if the frequency is f one obtains the relation*

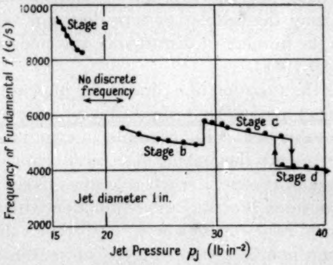


Fig. 12. Frequency of stream disturbance in axially symmetric case.

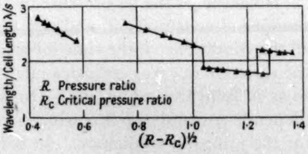


Fig. 13. Relation between fundamental radiated wavelength and cell length: axially symmetric case.

$$\frac{\theta}{f} = \frac{h}{\overline{M}c} + \frac{h - l\lambda}{c},$$

where $\overline{M}$ is the mean value. θ must take the form (integer + constant), say $N + p$, and is related to the number of disturbance wavelengths or, more crudely, eddies or vortices, in the flow.

Since for a given jet condition M, l and probably h are frequency dependent, it follows that there can only be a number of discrete solutions for (θ, f) according to the value of N since p will be fixed. The value of l, dependent on these, may change also at certain intervals.

---

* Identical with eqn. (3) of a former paper (Powell 1953) apart from a small change of notation. Also to eqn. (1), Powell 1952 a, if $l = 0$.

The other condition to be satisfied is that the 'gain' round the cycle of the motion cannot be less than unity, i.e. $q\eta_s\eta_t\eta_d \geqslant 1$, for the motion to become established. Here q denotes the amplification of the disturbance in the stream, η_s an efficiency by which it creates sound, η_t and η_d similarly referring to the transmission of sound to the orifice and the creation of a new stream disturbance. The interpretation of the pair of criteria above has been dealt with at some length in relation to edge tones in a previous paper (Powell 1952a).

In this application η_s and η_d will vary steadily with frequency, whilst q will have a single maximum. Thus, *so far as q is concerned*, the motion will take place only in a limited frequency range, so restricting the choice of value of (θ, f) to a few, perhaps one, or none at all at a limiting frequency.

But η_t is associated with the directionality of the sound, and so is acutely dependent upon (f, s) or (λ, s). We have seen in fact that the motion only exists over a certain range of λ/s close to 2·5 where η_t is near its maximum.* Hence the final criterion.

As the jet pressure increases, s, and therefore h, does likewise. The disturbance wavelength η must also increase (since $\lambda/s \simeq 2{\cdot}5$, and M does not vary much) for the same number of wavelengths to be present. If the phase condition can still be satisfied, and maintain η_t near the maximum (i.e. $\lambda/s \simeq 2{\cdot}5$), the motion will be continuous. Since the path of the cycle is longer, the frequency will fall steadily with increasing jet pressure. This is what occurs in the two-dimensional case, where the number of cells composing the cellular pattern does not vary greatly. It is significant, however, that the number of disturbance wavelengths present may vary for different jets (Powell 1953).

However, as the jet pressure increases the variation of h (due to the increasing number of cells present, and also their size) may make it impossible to maintain all the conditions all the time for a single value of θ. (For example, in edge tones the variation in q was critical, a jump in frequency then taking place as N changed, so as to fulfil the conditions.) In the case of the circular jet, it is most likely η_t (dependent mainly on λ/s) which is responsible, probably in conjunction with q, for the jumps in frequency. In contrast to edge tones (where $l = 0$†), the value of l may also vary by an integer at certain points, so that there is no reason to suggest that the changes in frequency will always be of the same nature as the jet pressure increases. This may well account for the different nature of the jumps occurring, as shown in fig. 12, perhaps surprising at first sight. If it so happens that other conditions cannot be satisfied, then no discrete frequency can exist, and the confused screeching is then attributable to motions commencing to build up, but becoming irregular and confused on violating the phase conditions (for example the region between stages 'a' and 'b', fig. 12). A possible causation for this phase violation is that h decreases as the amplitude decreases, due to the enhanced mixing.

Irregularities *within* stage 'd' had a different form, the note being of constant amplitude but intermittent at a frequency of a few cycles per second. An alternative explanation of this unsteadiness is that the plane of the motion was rotating about the jet axis (there being zeroes of course normal to the plane of symmetry) owing to a slight swirl in the flow. This did not arise, however, at the other stages.

* Other values of λ/s may be theoretically possible, but their mode is apparently inhibited by the one indicated (cf. edge tones).

† In the low speed case.

The High Frequency Sound

In addition to the waves considered above, others of a much higher frequency may be present, as in figs. 10 and 14 (Plate). The emission appears to be directional, being mainly within an angle of 60° to the jet direction. It appears to be associated with the boundary layer, which is laminar on leaving the exit. The short length of laminar flow can be seen in fig. 14 (for example) as a white line, black in fig. 6 and as a short collar lacking the granular appearance of the following turbulence in figs. 8 and 9, the latter being a subsonic case. Small-scale but intense disturbances can be seen to develop in the boundary layer near the end of this region (see especially figs. 7 and 14), and the associated density gradients extend a small distance into the external flow in a manner analogous to the larger disturbances. It is these disturbances, presumably originating from the break-up of the unstable boundary layer, from which the sound appears to emanate. Not all this sound emanates from this region, however, part of it appearing to emanate from further downstream, possibly from the shock positions. The phenomenon (first reported by Powell 1951) has also been noted by Lilley (1952), in which case the sound waves appear almost to take the form of shocks, their origin being rather more clearly seen to be as suggested above. The frequency of this type of emission is above the audio range, being in the region of 40–50 kc/s, although not taking the form of a simple note. (The range for a $\frac{1}{4}$ in. jet appears to be about 80–100 kc/s, Powell 1951.) This form of sound generated by a jet is clearly distinct from that responsible for the audible effects in which we are at present interested and will, therefore, not be considered further at the moment, although its presence is most interesting and should be borne in mind when considering other sizes of jets.

§ 5. CONCLUDING REMARKS

Apart from a high-frequency sound associated with the collapse of the initially laminar boundary layer, the outstanding characteristic of the sound of a choked jet is that a powerful screech of discrete frequency appears and swamps the noise due to turbulent mixing. The basic mechanism is similar in both the two-dimensional and axially symmetric cases. This is that the passage of an alternately disposed disturbance (or 'eddy') system gives rise to sound on traversing the regularly spaced shock wave system of the jet, the interference being such that a powerful emission takes place in the upstream direction. On passing the orifice the sound waves give rise to embryo disturbances, which become amplified as they pass downstream, ultimately to produce sound. In order that the process be maintained certain conditions of phase and gain must be fulfilled. These appear to be satisfied over the whole range investigated in the two-dimensional case by a single arrangement of the disturbance and sound wave system, the amplitude building up rapidly from zero at the lower limit as the gain criterion is fulfilled, but in the other case jumps in frequency and intensity (tending to produce 'steps') occur as the formation changes, i.e. as more eddies or sound waves appear between the orifice and the end of the cellular pattern, near which the effective source of sound is located. The latter is the most general case, and seems the most likely. There is nothing to suggest that the two-dimensional case is *always* continuous and the axially symmetric case *always* discontinuous.

In both cases the disturbance system is alternate, superficially like a Kármán Street, and the sound waves are consequently of opposite phase across the jet. In the case of the flow from a circular orifice some small eccentricity of the jet or perhaps acoustic reflections may determine the plane in which the alternate motion takes place, or the effects of swirl may cause this plane to rotate about the jet axis, distorting the motion. In both cases the ratio (sound wavelength)/(cell length) = 2·5, roughly, producing a powerful upstream radiation, and a high peak roughly normal to the stream at precisely double that frequency in the cases specificially examined.

The phenomenon has been placed on a qualitative basis: its mechanism has been explained and agreement with experiment is encouraging. Such properties like directionality can be predicted once certain characteristics of the flow are known. The scheme of a purely theoretical approach is apparent, based upon the solution of the phase equation

$$\frac{N+p}{f} = \int \frac{dh}{Mc} + \frac{h-l\lambda}{c},$$

together with the gain criterion $q\eta_s\eta_t\eta_d \geqslant 1$. The taks of solving these is very formidable. In the gain equation little is yet known either of the rates of amplification of the stream disturbances q in such a flow, particularly when the amplitude is large, or about the actual production of the sound by the disturbances interacting with the cellular pattern η_s. The other two factors are more amenable, the transmission of sound to the orifice η_t being associated with the directionality, which has already received some attention. The remaining factor η_d, the initiation of the stream disturbances, is associated with the pressure ratio across the jet boundary at the orifice, and since this fluctuates with the sound pressure there, the angle which the jet boundary takes up will also fluctuate, so giving rise to the initially small disturbance.

In the phase equation, h probably presents the greatest difficulty, being the distance of the effective source from the orifice. This is because the effective source is located near the end of the cellular pattern (where the disturbances are largest), the disintegration of which is likely to be dependent upon the amplitude of the strong periodic disturbances present as well as upon the general turbulence.

Thus a great step forward is necessary to place the mechanism on a purely theoretical basis, the amplitude of the motion probably appearing almost as a by-product of the determination of what frequency, if any, dominates. The use of empirical factors determined from experiment, for example, taking certain values for λ/s, or h, should be helpful in the earlier stages.

The noise emanating from a choked jet is of a complex nature: in addition to that generated in the manner discussed above there is also that produced by the general turbulence interacting with the shock waves. Further, there is the noise generated by the turbulent mixing, commencing with the break-up of the laminar boundary layer and developing into the subsonic eddying flow far downstream, just as in a subsonic jet. This type of noise will be enhanced in heated jets due to the higher jet velocities attainable. Thus although the 'screech' may completely predominate, as in the experiments described, it is not the only noise present, and in other jets, for example in the case of jet engine or rocket effluxes, the nature of the total noise will depend on the balance of these factors. The existence of the 'gain criterion' derived above gives a hint as to how the screeching noise night be reduced, i.e. by reducing one or more of the factors occurring in it, but it is clear that the total reductions in noise level will depend on the resultant balance between the types of noise produced by the various mechanisms. This important question of noise reduction will be considered in a future paper.

ACKNOWLEDGMENTS

The author wishes to express his sincere thanks to Professor E. J. Richards for his unfailing interest and constructive criticisms throughout the work, to Mr. K. R. McLachlan for the design and construction of the circuit for the photoelectric device, and to the Ministry of Supply, who have sponsored the work.

REFERENCES

BROWN, G. B., 1935, *Proc. Phys. Soc.*, **47,** 703.
HARROP, R., 1951, *J. Roy. Aero. Soc.*, **55,** 169.
LILLEY, G. M., 1952, 8th International Congress on Theoretical and Applied Mechanics, Istanbul.
POWELL, A., 1951, Report A.R.C. 14,726, F.M. 1694; 1952 a, Report A.R.C. 15,333, F.M. 1809; 1952 b, Report A.R.C. 15,473, F.M. 1837; 1953, *Aeronautical Quarterly*, **4,** 103.
SAVIC, P., 1941, *Phil. Mag.* **32,** 245.
WESTLEY, R., and LILLEY, G. M., 1952, *College of Aeronautics* Report No. 53.

APPENDIX

THE DYNAMIC SCHLIEREN APPARATUS

Notation

A	amplitude of simple acoustic source
a_n	constant in asymptotic series
b	$=\rho_0 K$
c	speed of sound $=(\gamma p_0/\rho_0)^{1/2}$
C	contrast of wave form on Schlieren photograph
E	width of undisturbed image at critical knife-edge
F	focal length of 'second' Schlieren mirror
f	frequency
G	sensitivity of the photoelectric device
I	intensity of sound, erg cm^{-2} sec^{-1}
K	Gladstone–Dale constant for air, $n-n_0=K(\rho-\rho_0)$
k	wave number $=2\pi/\lambda$
L	constant derived from asymptotic series
M	optical magnification of knife-edge movement
n	refractive index of air
p	increment of pressure, p_0 mean value
q	$=k(x^2+y^2)^{1/2}$
R	reading from photoelectric device
r_0	distance from source, $(x^2+y^2)^{1/2}$
s	amplitude of light shift at critical knife-edge
t	time
x	orthogonal coordinate of viewed field perpendicular to knife-edge
y	orthogonal coordinate of viewed field parallel to knife-edge
z	orthogonal coordinate of viewed field normal to plane of knife-edge
Δ	instrument reading
ϕ	velocity potential of sound field
γ	ratio of specific heats of air
λ	wavelength of sound wave
ρ	increment of air density, ρ_0 mean value
θ	$=\tan^{-1} y/x$

Basic Relationship

In many applications it is possible to obtain a reasonably accurate estimate of the sound energy progressing in any direction to which the optical system responds, provided the geometry of the sound field is known. Assuming this to be so, we have

$$\frac{\partial \rho}{\partial x}=\frac{1}{c^2}\frac{\partial p}{\partial x}=\frac{\rho_0}{c^2}\frac{\partial^2 \phi}{\partial x\,\partial t}.$$

Now a pencil of light passing through a field of variable density undergoes an angular deviation resulting in a linear deflection at the critical knife-edge of

$$s=F\int K\frac{\partial \rho}{\partial x}\,dz=\frac{Fb}{c^2}\int\frac{\partial^2\phi}{\partial x\,\partial t}\,dz,$$

where the integral is to be taken over the region where the integrand is non-zero. Hence if G is sensitivity of the photoelectric device and its indicator, the latter will indicate

$$R=\frac{GFb}{c^2}\int\frac{\partial^2\phi}{\partial x\,\partial t}\,dz. \qquad \text{......(1)}$$

This is the fundamental relationship. The problem in a given case is to find the nature of ϕ, its integral and the overall instrument sensitivity G.

Instrument Sensitivity

The instrument sensitivity is dependent upon several factors. So far as the light source is concerned, it is the illumination per unit width of the image at the critical knife-edge which matters, in conjunction with the undisturbed width of the image passing the knife-edge (owing to the nature of the response of the photomultiplier). The gain of the photomultiplier is acutely dependent upon the supply voltage which, if not efficiently stabilized, will vary with mains voltage, as will the power of the light source. These factors are variable and not amenable to calculation; a means of calibration which is quick and convenient linking linear movement at the knife-edge with instrument output must be used. A simple way of doing this is to mount the knife-edge on a driven tuning fork, of sufficiently high frequency to be within the region for which the overal response is flat (actually 512 c/s was used*). Its amplitude of vibration was measured by observing the deflection of a ray of light, focused upon a scale, from a tiny concave mirror of 1 m focal length attached to the tip of the fork. This produced a magnification M of the knife-edge movement of 34 in the example given.

Thus if the fork had an amplitude of vibration s', the instrument response would be $R'=Gs'$ and the scale reading $\Delta'=Ms'$. Then for an instrument reading R the shift at the knife-edge is simply $s=R\Delta'/MR'$. Actually the indicator used was a cathode-ray oscilloscope.

An Application

One application of the method is in finding the sound intensity in the field of the two-dimensional jet. Schlieren photographs (see Powell 1953) have shown that the sound effectively emanates from an equivalent line source parallel to the light paths. At a point several wavelengths away, in a plane lying within the boundary planes of the jet, the acoustic motions will closely resemble those of a simple source of certain strength at the point of origin. If the amplitude

* A higher frequency is really desirable for this work

of this equivalent source is A, then the acoustic power transmitted across unit area normal to the vector from it will be

$$I=\rho_0 c\,(k^2A^2/2r_0^2) \qquad (z=0) \qquad \ldots\ldots(2)$$

and the potential function is

$$\phi=\frac{A}{r}\,\mathrm{e}^{ik(ct-r)}. \qquad \ldots\ldots(3)$$

The solution to the integral of eqn. (1), which can be found by contour integration, becomes, for kr moderately large,

$$A\frac{\partial^2}{\partial x\,\partial t}\int\frac{\mathrm{e}^{ik(ct-r)}}{r}\,dz=A\frac{\partial^2}{\partial x\,\partial t}\mathrm{e}^{ikct}\left(\frac{\pi}{2iq}\right)^{1/2}\mathrm{e}^{-iq}\,\Sigma_0 a_n q^{-n}, \qquad q^2=k^2(x^2+y^2),$$

where $a_0=1$, $a_1=\dfrac{1^2}{1!}\left(\dfrac{i}{8}\right)$, $a_2=\dfrac{1^2.\,3^2}{3!}\left(\dfrac{i}{8}\right)^2$ etc. Let $a_0=a_0'$, $a_1=ia_1'$, $a_2=-a_2'$ etc.

The series is asymptotic, the terms at first decreasing and then after a certain point increasing indefinitely. If the process of summing is stopped at a small term, the error will be small (for q not too small). The series may be differentiated provided this does not change its nature. The preceding expression then becomes

$$A\left(\frac{\pi}{2}\right)^{1/2}k^3cxq^{-3/2}L\cos\left(kct-q-\tfrac{1}{4}\pi-l\right), \qquad \ldots\ldots(4)$$

where $L^2=[a_0'-(a_2'-3a_1'/2)q^{-2}+(a_4'-7a_3'/2)q^{-4}\ldots]^2$

$\qquad +[-(a_1'-a_0'/2)q^{-1}+(a_3'-5a_2'/2)q^{-3}\ldots]^2$

and $\quad l=\tan^{-1}[a_0'-\ldots]/[-(a_1'-a_0'/2)q^{-1}+\ldots]$.

The shift of the light beam will be Fb/c^2 times expression (4). The value of $L=1$ will be a close approximation for the range of q in which we are interested.

Hence for a given instrument reading the intensity of the source at the point considered will be, from the preceding expression and eqns. (1) and (2),

$$I=\rho_0 c\,\frac{c^3}{2\pi^2}\,\frac{1}{fr_0\cos^2\theta}\left(\frac{R}{LGFK}\right)^2, \qquad \ldots\ldots(5)$$

where of course $R/G=s=R\Delta'/MR'$, the primes indicating calibration values.

As an example of the magnitudes found in this work, for a certain case, $f=20\,000$ c/s, $r=5{\cdot}00$ cm, $\theta=30°$ and $V=20$ v. Calibration: $R'=18$ v corresponded to scale deflection $\Delta'=0{\cdot}15$ cm with $M=34$. Thus $s=0{\cdot}004\,90$ cm. Since $F=183$ cm, $K=0{\cdot}000\,294$, and since here $L=1$, very nearly, $I=9{\cdot}25\times10^6$ erg cm^{-2} sec^{-1}, or 160 db, very nearly, above the usual reference pressure of $0{\cdot}000\,2$ dyn cm^{-2}.

It is interesting to note that had plane sound waves been considered a result similar to (5) would have been arrived at, but greater by a factor $\lambda r_0/4n^2$, where λ is the wavelength and n the depth of the field. The two forms are evidently equivalent if $n=\frac{1}{2}(\lambda r_0)^{1/2}$.

The shift could be also obtained approximately directly from instantaneous Schlieren photographs provided refraction effects are negligible, for if the contrast is C and the undisturbed image width E,

$$C=\frac{E+s}{E-s} \quad \text{and so} \quad s=E\,\frac{C-1}{C+1}. \qquad \ldots\ldots(6)$$

Perhaps the most satisfactory way of finding C would be to insert a series of known gradients into the field for comparative purposes or for the making of slides under controlled conditions.

Comments on the Method

It has been shown how the method can be used to find sound intensity provided the geometry of the field is known and the associated density gradients are sufficiently high. The latter is necessary to ensure a satisfactory signal-to-noise ratio since a d.c. component of the input is inevitable and gives rise to a certain electronic 'noise'. It may of course be kept to a minimum by increasing the cut-off to its limit, any further increase then resulting in flat-topped wave forms.

The great advantage of the method is that there is no interference whatsoever with the sound field, and the increasing sensitivity with frequency, contrary to many high frequency microphones, is a useful feature in certain applications.

The density gradients need not be acoustic: they may be those associated with turbulence. The method can be directly applied to certain unsteady flow problems, the frequency range far exceeding that of hot wire techniques, the measured quantity of course being quite different. The absence of any interference with the flow is a valuable asset, particularly where the stability may be of a critical nature, as in the present application. If the structure of the disturbance system could be considered to be two-dimensional, a spectrum of density gradient in any given direction (or, with modification, directions) can be immediately obtained by analysis of the output wave form.

The values of such measurments of a turbulent flow (or derived quantities) have yet to be assessed. They are likely to be of particular interest in connection with the estimation of sound generated by a turbulent flow, or the mixing of fluids of different refractive index (i.e. perhaps of different density or temperature).

Certain correlations may be obtained most simply. By using two apertures the semi-wavelength of the eddy system was found, simply by finding the aperture spacing corresponding to the removal of the fundamental and odd harmonics and doubling of the even harmonics. (The method can be simply adapted for the case where the wave form changes appreciably over the semi-wavelength.) Since the frequency is known, the translational velocity of the eddy system follows immediately. Clearly by using a narrow-band filter this might be done for any component of the spectrum. The two-dimensional nature of the method is an unfortunate limitation, inherent in the optical system, unless some statistical means of unscrambling the averaged signal can be found. It is hoped to examine the possibilities of modifying this to enable application of the method to important three-dimensional turbulent flows. Of course such properties as frequency can be determined in the three-dimensional case if there is a correlation over a significant depth of the field, for example as in the case of the disturbance processions in the round jet.

The Penetration of Electrons into Luminescent Material*

By W. EHRENBERG and J. FRANKS†
Department of Physics, Birkbeck College, London

MS. received 19th May 1953, and in amended form 10th August 1953

Abstract. The penetration of electrons of energies between 10 and 40 kev into a number of phosphors, and the extent of their scattering, is measured on microphotographs of the luminous region produced in the phosphor by a beam of electrons of less than 0·75 μ diameter. A description is given of the apparatus and of the electron optical system designed to produce the fine focal spot. The phosphors used in the experiment were single crystals of the thallium-activated iodides of K, Rb and Cs, and of the tungstates of Ca and Cd, and a luminescent plastic. An attempt is made to account for the discrepancies in the electron ranges as found in the present experiments compared with those obtained by other methods. An estimate is made of the distribution in depth of the energy dissipation in a phosphor when bombarded by a beam of electrons.

§ 1. INTRODUCTION

A KNOWLEDGE of the penetration and scattering of electrons in solids is essential in the study of phenomena related to the interaction of electrons with matter, such as luminescence and secondary emission. The penetration of electrons of high energies (2–6 Mev) in a KCl crystal was studied by Berger and Paul (1949) by measurement of the colour centres produced; the interpretation of the results was, however, complicated by the recombination of the centres. No direct observations of penetration and scattering in solids of electrons with incident energies of some tens of kev have so far been published, although recent work on luminescent single crystals (Taylor *et al.* 1951) and thin phosphor films (Koller and Alden 1951) enabled a certain amount to be inferred about the behaviour of electrons in transparent luminescent solids. The results presented here are based on inspection by microscope of the luminous figure produced by a narrow bundle of electrons incident on a phosphor.

This luminous figure, for brevity called glow, represents an average of the single luminous electron tracks, and as such provides information about the limits of penetration and the scatter of electrons, and the distribution in space of the dissipation of their energy.

§ 2. EXPERIMENTAL ARRANGEMENT

The crystals under test were situated at the bottom of a long vertical cathode-ray tube, so that the electrons entered a flat horizontal face and the glow was observed through an adjacent flat vertical face by means of a horizontal microscope. In order to give the assembly the required rigidity and freedom from vibration it was clamped to a strong U-shaped cast aluminium girder provided with a platform, on which the microscope rested. The greater part of the cathode-ray tube was screened by mumetal 0·004 in. thick, in order to suppress the disturbing influence of alternating fields of 50 c/s and harmonics originating from mains supplies in the building. The shielding did not entirely prevent static deflection of the beam by the earth's horizontal field, necessitating the use of degaussing coils.

---

* This paper is based on a thesis by J. Franks which has been approved by the University of London for the degree of Ph.D.

† Now at Associated Electrical Industries, Aldermaston, Berks.

The observations were made with an electron beam of a width near the limit of microscopic resolution; no useful purpose would have been served by going below that, while a wider beam would have interfered with the interpretation of the observed glows. The fine beam was produced by demagnifying the virtual source in the neighbourhood of a filamentary electron gun, by means of a strong magnetic lens. The electron gun permitted a hairpin filament to be displaced longitudinally and vertically with respect to a 4 mm aperture in the surrounding Wehnelt cylinder which projected into an earthed 2 in. diameter tube forming the anode. With a voltage of 10–40 kv (derived from a stabilized radio-frequency unit) applied to the cathode and a bias of about −50 v applied to the Wehnelt cylinder, a very fine electron source was obtained when the tip of the filament was about 6 mm behind the aperture of the cylinder. Under these conditions the intensity of the beam at the crystal had dropped considerably from its maximum near the cut-off potential but nevertheless remained ample for observation.

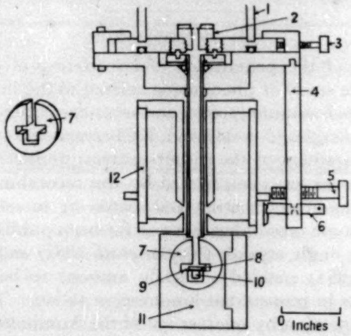


Fig. 1. The sliding plates, lens adjustment and crystal chamber.
1, 30 in. tube; 2, lens aperture stop; 3, three adjusting screws at 120°; 4, three lens adjustment sets at 120° (square cross section); 5, lens adjustment screw (horizontal displacement); 6, lens adjustment cam (tilt); 7, crystal holder; 8, crystal under test; 9, glass plate; 10, glass prism; 11, crystal chamber; 12, magnetic lens (winding not shown).

As the focal length of the magnetic lens could not conveniently be reduced to much below 1 in., the anode tube was made 30 in. long in order to obtain a demagnification of about 40. The magnetic lens was of conventional design, with a carefully machined and annealed 'Permendur' shield; with a 0·1 mm stop it produced an image of the fine source of less than 0·75 μ diameter, carrying a current estimated to 10^{-11} A.

The bottom portion of the cathode-ray tube, including the supports for the magnetic lens and the specimen chamber, is shown in fig. 1. The construction of the chamber was governed by the necessity of placing the bombarded portion of the specimen *in vacuo* within working distance of the (high power) microscope objective and also immediately below the magnetic lens. The axis of the narrow tube from which the specimen chamber was suspended passed $\frac{1}{32}$ in. behind the window of the chamber (a 0·18 mm microscope cover-slip); the narrow tube could be horizontally displaced with respect to the main tube, and the magnetic lens could be adjusted with respect to the narrow tube. Different parts of the specimen could thus be exposed to the beam within moderate limits. The specimen itself rested on a thin glass plate, to which a 90° prism was attached. This made it possible to obtain an axial view of the illuminated region of the specimen with low magnification, while the main observations were made by viewing the crystal in a direction normal to the electron beam.

The choice of the microscope objective required careful attention. Commercial objectives are either corrected for a cover-slip thickness of 0·18 mm or are designed for use with uncovered objects. When observations are made through an additional thickness of dielectric, a negative spherical aberration is introduced. To obtain an aberration-free image of the glow inside the phosphor a 0·65 N.A. 40 × objective was used, in which the spacing between the front components was reduced so that it was free from spherical aberration for a cover-slip thickness of 0·4 mm. The objective was thus corrected for one fixed depth only, which depended on the refractive index of the specimen. It was therefore necessary to adjust each time successively the location of the illuminated region, the magnetic focus and the optical focus, repeating a cycle of operations until an image free from aberration was obtained. Visual observations were made with a 15 × Kellner eyepiece. To photograph the glow a special camera was inserted between the microscope tube and the eyepiece, which permitted replacing the eyepeice scale by successive sections of a film 2 in. long and $\frac{1}{2}$ in. wide. A series of seven photographs could be taken on one film.

§3. PREPARATION OF SPECIMENS

Measurements were made on specimens polished or cleaved into rectangular prisms. Both the front face, placed flush against the window of the specimen chamber, and the adjoining upper face, exposed squarely to the beam, were required to be of optical perfection to avoid distortion of the image and light scatter. This applied especially to the region on either face close to the common bounding edge, which had to be sharp and flawless because the depth of penetration of the electrons never totalled more than a few microns. Further, a large proportion of the light produced in the luminescing region was totally reflected by the upper surface, producing a mirror image of the glow of an intensity almost equal to that of the glow itself. Owing to the small depth of focus of high power objectives, serious blurring occurred when these two objects were not in the same plane: the front and upper faces were thus required to be accurately normal to each other. Finally, the face resting on the glass plate had to be parallel to the upper face and of fair optical flatness, and the specimen as a whole had to be transparent and clear, so that the upper face could readily be observed with a low power objective through the specimen and prism.

Clear single crystals of RbI(Tl) and KI(Tl) grown from the melt by a modified Stöber technique (Stöber 1924, Franks 1953) could readily be cleaved into the required shape. The cleavage planes of these crystals were generally bounded by sharp edges, but the faces usually exhibited lineage to a greater or lesser extent. Occasionally, however, the cleaved crystals were almost flawless near one edge and could be used without further preparation. Of the other thallium-activated alkali halide crystals grown, only NaI(Tl) and CsI(Tl) gave a sufficiently intense luminescence in the visible region. The NaI(Tl) crystals, although easily cleaved, were unsuitable: a special technique would have been required to handle this very hygroscopic substance. The CsI(Tl) crystal could not be cleaved and was cut and polished by Messrs. Hilger and Watts. This firm also polished crystals of $CaWO_4$ and $CdWO_4$ obtained from Linde Air Products, Division of Union Carbide and Carbon Corporation, Chicago, Illinois, and a fluorescent plastic sample obtained from the French Atomic Energy Commission.

The upper face of several of the specimens was coated with a thin film of gold in a vacuum evaporating chamber to enable the surface to be earthed during bombardment.

§4. RESULTS

The luminescent regions produced by electrons of 10, 20, 30 and 40 kev in the specimens were examined both visually and by photographs. At every voltage a series of exposures of 10, 20, 40, 80 and 120 sec was taken on one film, the tube current having been adjusted to such a value that the 10 sec photograph was slightly underexposed, while the longer exposed photographs tended to become identical in blackening and features. By this procedure the glow could readily be distinguished from the background and from the spurious blackening round the image, both of which were attributable to the scattering and reflection of light in the crystal and halation in the film—an important factor when the image extends over relatively few grains. The contributions from these disturbing effects became, of course, most marked in long exposures, after the blackening due to the true glow had reached saturation.

The glow remained steady and symmetrical with all the phosphors except the plastic while observations were made, the glass window in close proximity to the beam apparently causing no disturbance due to electrostatic charging. Furthermore, none of the crystal phosphors exhibited any 'sticking potential' effect, the depth of penetration steadily increasing with the voltage. This was attributed to slower secondary electrons from adjacent metal walls, produced by fast secondaries from the phosphors, which maintained the insulators at earth potential. To confirm that the specimens were observed under earthed surface conditions, the upper faces of the tungstate crystals were coated with a thin film of gold after a set of observations had been made with uncoated faces. The coated faces were earthed by packing gold foil between the film and the top of the horizontal cavity in the specimen holder. No systematic difference was detected between the results obtained without a gold-coated surface and those obtained with the surface earthed. The surface of the uncoated luminescent plastic became charged however, as was evident from the initial rapid shrinking of the penetration figure to a small luminescing region independent of the incident voltage. All measurements with this phosphor were therefore made with an earthed gold-coated surface, under which condition consistent results were obtained.

The behaviour of the plastic (density $1{\cdot}05\,\mathrm{g\,cm^{-3}}$) may have been due to its small back-diffusion coefficient; this quantity decreases with the density of the bombarded substance (Schonland 1923, 1925, Palluel 1947). The number of electrons diffusing back to the surface was in this case apparently insufficient to produce the required amount of secondaries from the metal walls to maintain the surface of the plastic in an uncharged state.

The alkali halides were not tested with an earthed gold surface, as evidently the secondary emission was in all cases sufficient to prevent sticking.

A selection of typical photographic records is shown in fig. 2 (Plate). All show the glow itself together with its mirror image, as explained before; the mirror image is fainter only in the case of the fluorescent plastic. For substances of high density (such as $CdWO_4$, density $7{\cdot}9\,g\,cm^{-3}$) and at low voltage the glow resembles a hemisphere which, together with its mirror image, produces a circular disc on the photograph. This suggests that a large number of high energy secondaries are emitted from substances of higher density, while the mushroom-shaped figure obtained with the plastic shows that almost every electron loses all its energy within this phosphor. Thus the records lend colour to Schonland's and Palluel's measurements of back diffusion of electrons.

The photographic records show moderately well defined contours, enlargements of which were drawn on paper by means of a *camera lucida*. The contours denote lines of equal luminosity in the projected glow. For very long exposures these are distorted by scattered light. The central region hardly ever increases in size by more than 30%. The influence of scattered light on the records was checked by photographs of known test objects such as pinholes in metallic deposits.

Within one series luminosity at the contours is about inversely proportional to the time of exposure. In fig. 3 a set of two or three contours is shown for each material and voltage examined; it appears that the projected glows have a core of almost constant luminosity, and the luminosity drops rapidly to a negligible value.

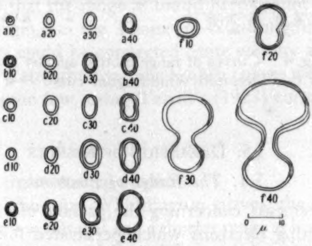


Fig. 3. Contours of glow photographs.
(*a*) $CdWO_4$; (*b*) $CaWO_4$; (*c*) CsI; (*d*) RbI; (*e*) KI; (*f*) plastic phosphor.
The figures denote the beam voltage in kv.

The contours are shown for an electron beam travelling from the bottom of the page to the top. A (100) face of the cubic RbI and KI crystals was exposed to the beam, a (110) face of the tetragonal $CaWO_4$ and a (010) face of the monoclinic $CdWO_4$. The inner contours are those of the 10 sec exposures, the outer contours represent exposure of 80 sec for 40 kv figures; in all other cases an exposure of 120 sec; intermediate lines correspond to a 40 sec exposure.

The luminosity in turn can be taken as a measure of the density of dissipation of energy in the phosphor if allowance is made for the fact that the observed luminosity is the projection of an axially symmetric glow.

In rough approximation for the plastic phosphor, and rather well in the other cases, the contours resemble sections of a circle, the centre of which is near the surface for low voltage and dense materials and further below the surface for higher voltages and lighter materials. This suggests that the loss of direction an electron beam suffers in solid dielectrics becomes catastrophic almost as soon as it commences; it commences in dense solids at the instant of entry for all voltages here investigated.

Further, the 'practical' range of electrons R should be given by the contour line extrapolated for infinite exposure, as the distance between the point of entry of electrons to the point of the contour opposite to it. Graphs showing this range plotted against voltage for various substances under test are given in fig. 4. The points can equally well be connected by a straight line or by a parabola.

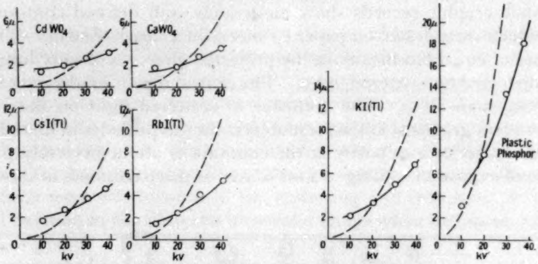


Fig. 4. Curves of range plotted against voltage.

The broken line is the Thomson-Whiddington curve $r=V^2/b$. The full line is the curve $r=V^2/c+d$.

§ 5. Discussion of Results

5.1. *The Range of Electrons*

Hitherto, measurements concerning the passage of electrons through solids were made by detecting electrons which penetrated foils (usually metallic) of increasing thickness, while in the experiments described here the electrons could not entirely traverse the bombarded material. In this respect the present arrangement reproduces conditions in cloud chamber range determinations (Williams 1931, Tsien 1944).

In the foil experiments either the energy or the energy spectrum of the transmitted electrons is measured by magnetic deflection methods (Whiddington 1912, Terrill 1923, White and Millington 1928), or the number of transmitted electrons is measured and absorption curves obtained. In the latter type of observation the charge of the electrons may be measured with a Faraday cylinder or Geiger counter, thus giving their number (Whiddington 1914, Terrill 1924, Schonland 1923, 1925, Eddy 1928), or an ionization gauge may be used as detecting device, in which case the result obtained depends both on the number and energy of the electrons (Varder 1915, Madgwick 1927). This is also the case here, as the intensity of luminescence depends both on the number of electrons and on their rate of loss of energy. Bothe (1933) states that the ionization absorption curves do not differ appreciably in character from those obtained with the charge method.

It is most remarkable that, in spite of the diversity of the influence of scattering in the different techniques, the results, obtained with both the foil and cloud chamber methods for electrons of velocities considered in this paper, invariably agree with the Thomson–Whiddington law (Whiddington 1912), $v_0^4 - v^4 = ax$, where v_0 is the velocity of the electrons incident on a foil or the initial velocity of the electrons produced in a cloud chamber, v the velocity after the electron has travelled a distance x measured along its path, and a is a constant depending on the medium. The law can also be stated in terms of the potentials V_0, V, corresponding to the velocities v_0, v, thus

$$V_0^2 - V^2 = bx. \qquad \ldots\ldots(1)$$

Theoretical derivations of this expression, or close approximations to it, were obtained by Thomson (1906) and later by Bohr (1913, 1915) from classical considerations, and by Bethe (1930) from a quantum mechanical treatment. The theories refer to the 'true' range, i.e. the integrated path length or the path in the absence of scattering. Why this law should hold under the actual experimental conditions has never been explained. But the experimental evidence in its favour is very strong.

The range of electrons in matter (obtained by setting $V = 0$) is therefore proportional to the square of the energy of the incident electrons, and this is again in close agreement with experimental results previously obtained. The present measurements appear, on the contrary, to indicate that an almost linear relationship exists between the energy and range of electrons.

It is seen from fig. 4 that the range as found here cannot be represented by a parabola through the origin, i.e. the Thomson–Whiddington law is not obeyed. The experimental points could be connected quite well by a parabola not going through the origin or by a straight line; the broken curves were constructed from the Thomson–Whiddington law, using Terrill's (1923) constant (relating to the most probable energy)

$$b = 4 \times 10^{11} \rho \, \mathrm{v}^2 \, \mathrm{cm}^{-1}$$

where ρ is the density of the scattering medium.

Compared with the Thomson–Whiddington curves, the experimental points indicate a longer range at low voltages and a shorter range at higher voltages. This discrepancy is not readily corrected by a different value of b, and is outside the margin of error as the difference in size between the 10 and 20 kv glows was easily observed both visually and on the photographs.

The reason for the deviation of the range curves presented here from previous observations, and hence from the Thomson–Whiddington law, is not obvious. It was found impossible to reconcile the present results with previous work or to propose a theory on the basis of a linear relationship between voltage and range. A parabolic relation of the form of the Thomson–Whiddington law, but not passing through the origin, appeared more promising. This may be explained by the assumption that energy is transferred beyond the region to which the primary electrons are confined, by agents such as excitons, diffusion of slow secondary electrons or ultra-violet radiation, if this energy transfer takes place over distances of the order of one micron. The zero point may of course lie on the true range curve, but as long as the bombarding electrons have sufficient energy to excite luminescence in the phosphor, emission would take place, according to the above view, over the whole region bounded by the energy transfer range d. Thus even if measurements of range against voltage were made down to a voltage corresponding to the threshold excitation energy, the range extrapolated to zero energy would not pass through the origin but would intersect the range axis at a height corresponding to d. But the higher the voltage, the less important should be its contribution to the observed range. The total range may then be given to a first approximation by a modified Thomson–Whiddington range relation

$$R = V_0^2/c + d \qquad \ldots\ldots(2)$$

where V_0 is the incident voltage and R the observed apparent range. c and d were obtained from the experimental curve for each of the substances investigated and their values are given in the table. c is larger than Terrill's b, but c/b remains fairly constant, i.e. c is approximately proportional to the density. In comparing c and b it should however be remembered that V_0^2/b represents the 'true' range, whereas R is the depth beyond which no luminosity comparable with the body of the glow is observed and is taken as denoting the boundary between appreciable and negligible energy dissipation.

5.2. *The Dissipation of Energy*

The observed distribution of light may be compared with that due to a luminosity uniformly filling the major segment of a sphere, the surface of the dielectric being the chordal plane. In projection this would look like a luminous segment of a circle with an intensity falling off as $(1-y^2)^{1/2}$, where y is the distance from the centre expressed as a fraction of the radius; i.e. with an intensity which is almost constant over a large central area, the main drop occurring within one-third of the radius from the edge. Such a distribution agrees sufficiently well with the observed contours to permit the conclusion that in first approximation the energy of a narrow bundle of electrons of velocities here examined is dissipated uniformly within the major segment of a sphere. The range R is then the sum of the radius r of the sphere and the depth x_0 of its centre below the surface, or, with $\alpha = x_0/r$,

$$R = r(1+\alpha). \qquad \ldots\ldots(3)$$

These quantities were found from the tracings, fig. 3; values for α are given in the last four columns of the table.

	Density	$b \times 10^{-11}$	$c \times 10^{-11}$	d		α at			
Phosphor	(g cm^{-3})	(v^2 cm^{-1})	(v^2 cm^{-1})	(μ)	c/b	10	20	30	40 kv
$CdWO_4$	7·9	31·6	100	1·6	3·2	0	0·1	0·2	0·3
$CaWO_4$	6·12	24·5	71	1·3	2·9	—	—	0·1	0·2
CsI	4·51	18·0	52	1·9	2·9	0	0·3	0·5	0·7
RbI	3·55	14·2	41	1·5	2·9	0·2	0·4	0·4	0·6
KI	3·13	12·5	30	1·9	2·4	0·3	0·3	0·5	0·5
Plastic	1·05	4·2	9·8	3·7	2·3	—	0·4	0·8	0·8

It is also of interest to know the total energy W dissipated between the surface and a given depth x, and hence the energy dW dissipated in any thin layer parallel to the surface. If the incident energy is W_0, the ratios W/W_0 and dW/W_0 are independent of the width of the beam because wide beams can, of course, be considered as made up of narrow beams with overlapping glows. In the approximation discussed above, W is equal to the volume of a segment of a sphere of radius r, bounded by planes situated at distance x_0 and $x-x_0$ from the centre, multiplied by the density of energy dissipation which is given as the ratio of W_0 and the volume of the glow. Hence by elementary geometry

$$\frac{W}{W_0}=\frac{\beta(3-3\alpha^2+3\beta\alpha-\beta^2)}{2-\alpha^3+3\alpha} \qquad \dots\dots(4)$$

where $\beta=x/r$. The energy dissipated within a thin layer dx comes to

$$dW=\frac{3\{1-(\beta-\alpha)^2\}W_0\,dx}{r(2-\alpha^3+3\alpha)} \qquad \dots\dots(5)$$

which has a maximum at $\alpha=\beta$ or $x=x_0$; the energy dissipated per unit thickness increases at first and then decreases monotonically to zero at $x=R$. The initial increase, however, is slight for moderate values of α. For $CdWO_4$, for example at 30 kv, $\alpha=0{\cdot}2$, the differential dissipation at the surface is 96% of that at the maximum. If I is the incident current (5) can, with (2) and (3), be written

$$dW=\frac{3IV_0\,dx}{V_0^2/c+d}\,\frac{1-(\beta-\alpha)^2}{2+\alpha-\alpha^2}. \qquad \dots\dots(6)$$

At the surface $\beta=0$, and for small values of V_0, α is small and $d \gg V_0^2/c$, so that the energy dissipated in a surface film rises proportionally to V_0. For large values of V_0, d becomes small compared with V_0^2/c, and α varies only slowly with voltage, so that the dissipation decreases roughly inversely proportional to the voltage. Such effects have in fact been observed (cf. Koller and Alden 1951).

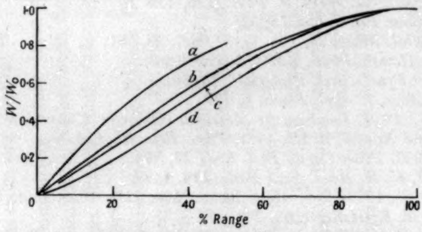


Fig. 5. Energy loss of electron beams: *a*, according to Stinchfield; *b*, *c*, *d*, according to eqn. (4) with $\alpha=0$, $\alpha=0{\cdot}33$, $\alpha=0{\cdot}70$ respectively.

Stinchfield (1940, quoted by Garlick 1949, p. 194) and others have proposed to calculate the dissipation of energy of electrons on the basis of a combination of the Thomson–Whiddington law (as confirmed by Terrill (1923) for the most probable energy of electrons leaving a foil in the direction of the primary beam and having lost less than half of the original energy) and of Lenard's law $N=N_0C^{-\gamma x}$ for the decrease of the electron current N with x. The fraction of the original energy remaining in the beam after the electrons have passed x is taken as NV/N_0V_0, so that $W/W_0=1-NV/N_0V_0$. For γ in Lenard's law Stinchfield further follows Terrill (1924) by equating $\gamma=k\rho/(v/2+v_0/2)^4$, i.e. setting γ proportional to the density of the material and inversely proportional to the fourth power of the mean of the initial and the most probable final velocity. Terrill's γ refers to the totality of the electrons leaving a foil with the exception of back-scattered and slow secondary electrons. W/W_0 then becomes a function of x/R. In fig. 5 Stinchfield's curve for W/W_0 is shown together with curves computed from eqn. (4) with $\alpha=0$, $\alpha=0{\cdot}33$ and $\alpha=0{\cdot}7$. These curves are very similar in spite of the difference in origin and in the algebraic expression. The energy dissipated is initially almost proportional to the thickness, and a large fraction is dissipated in the first half of the range. It should however be remembered that only in the neighbourhood of 20 kv the percentage range denotes the same depth in the medium, the range for (*a*) being given by eqn. (1) or the broken lines of fig. 4 and the range for (*b*)–(*d*) by eqn. (2) or the full lines of fig. 4. For the depth below the surface as abscissa the difference between (*a*) and the other curves would decrease at higher voltages and increase at lower voltages.

Stinchfield's case is weak, however, because Lenard's law is weak and because the dominating effects of scattering and straggling are not explicitly taken into account, though they are to some extent allowed for by the laws he uses. Equation (5), on the other hand, is based on a rather ruthless simplification of experimental material of moderate accuracy: still, it is probably the best 'law' available.

ACKNOWLEDGMENTS

The work described in this paper was carried out in the Physics Department of Birkbeck College, University of London. One of the authors (J. F.) was in receipt of a maintenance grant from the Department of Scientific and Industrial Research.

REFERENCES

BERGER, H., and PAUL, W., 1949, *Z. Phys.*, **126,** 422.
BETHE, A., 1930, *Ann. Phys., Lpz.*, **5,** 325.
BOHR, N., 1913, *Phil. Mag.*, **25,** 10; 1915, *Ibid.*, **30,** 581.
BOTHE, W., 1933, *Handb. Phys.*, XXII, **2** (2nd Edn.).
EDDY, C. E., 1928, *Proc. Camb. Phil. Soc.*, **25,** 50.
FRANKS, J., 1953, *Brit. J. Appl. Phys.*, **4,** 377.
GARLICK, G. F. J., 1949, *Luminescent Materials* (Oxford: Clarendon Press).
KOLLER, L. R., and ALDEN, E. D., 1951, *Phys. Rev.*, **83,** 684.
MADGWICK, E., 1927, *Proc. Camb. Phil. Soc.*, **23,** 970.
PALLUEL, P., 1947, *C. R. Acad. Sci., Paris*, **224,** 1492.
SCHONLAND, B. F. J., 1923, *Proc. Roy. Soc.* A, **104,** 235; 1925, *Ibid.*, **108,** 187.
STÖBER, F., 1924, *Z. Kristallogr.*, **61,** 299.
TAYLOR, C. J., REMLEY, M. E., JENTSCHKE, W. K., and KRUGER, P. G., 1951, *Phys. Rev.*, **83,** 170.
TERRILL, H. M., 1923, *Phys. Rev.*, **22,** 101; 1924, *Ibid.*, **24,** 616.
THOMSON, J. J., 1906, *Conduction of Electricity through Gases* (Cambridge: University Press), pp. 375 etc.
TSIEN, S. T., 1944, *Ann. Phys., Paris*, **19,** 327.
VARDER, R. W., 1915, *Phil. Mag.*, **29,** 725.
WHIDDINGTON, R., 1912, *Proc. Roy. Soc.* A, **86,** 360; 1914, *Ibid.*, **89,** 554.
WHITE, P., and MILLINGTON, G., 1928, *Proc. Roy. Soc.* A, **120,** 701.
WILLIAMS, E. J., 1931, *Proc. Roy. Soc.* A, **130,** 310.

Water Bells

By G. N. LANCE and R. L. PERRY
Department of Mathematics, King's College, London

Communicated by F. L. Hopwood; MS. received 31*st August* 1953

Abstract. The authors have studied theoretically the water bell phenomena, interest in which was revived recently by Hopwood. A numerical method of solution of the differential equation of the bell's surface is described and the results obtained are given in the form of graphs. It is shown that the shapes depend on three parameters and, in order to determine numerical magnitudes of these, experiments similar to Hopwood's were performed, during the course of which some photographs of the bells were obtained. Four of these are included. The paper ends with a section comparing the theoretical shapes with those obtained experimentally showing the agreement between the two to be good.

---

§ 1. INTRODUCTION

The subject of water bells is by no means a modern one. The earliest reference to them is in papers by Savart (1833, 1834); a further study was made by Boussinesq (1869, 1913) and in later days the bells have been used to determine the surface tensions of various liquids. The subject was revived recently by Hopwood (1952) who described a new method of producing the bells and gave a description of the novel experimental results he obtained.

The present paper gives a theoretical treatment of the subject as well as describing some further experimental work. In the theoretical section it is shown that the shapes of the bells depend on three parameters; in order to get solutions of the equation involved it is necessary to know the numerical magnitude of these parameters. Hence it was decided to repeat the experiments in a slightly different way from that used by Hopwood.

§ 2. DESCRIPTION OF THE APPARATUS AND OF THE EXPERIMENTAL METHOD

The apparatus, shown in fig. 1, consists of a hollow brass tube to the top of which is fitted a cap in such a way that there is a slot between the cap and the top of the tube. The width of the slot can be varied from zero to a maximum value of 0·1 in. determined by the size of the tube.

There is a perforated disc about 1½ in. from the top cap, the idea of which is to make the water project horizontally and steadily from the tube. The tube whose diameter is 1 in. is placed vertically in a trough of water so that the slot is about 2 in. above the water surface.

The experiment is performed in the following way. Water is passed through the tube, and if the bell so formed is punctured with the finger—thus equating air pressures inside and out—the shape shown in plate I is obtained. In the notation introduced below this is equivalent to putting $\alpha = 0$. The water flow and the size of the slot are now left unchanged but air bubbles are blown into the bell, without piercing the shell, thus making α negative. The bell changes shape the initial and the most probable final velocity. Terrill's γ refers to the totality of the electrons leaving a foil with the exception of back-scattered and slow secondary electrons. W/W_0 then becomes a function of x/R. In fig. 5 Stinchfield's curve for W/W_0 is shown together with curves computed from eqn. (4) with $\alpha = 0$, $\alpha = 0{\cdot}33$ and $\alpha = 0{\cdot}7$. These curves are very similar in spite of the difference in origin and in the algebraic expression. The energy dissipated is initially almost proportional to the thickness, and a large fraction is dissipated in the first half of the range. It should however be remembered that only in the neighbourhood of 20 kv the percentage range denotes the same depth in the medium, the range for (*a*) being given by eqn. (1) or the broken lines of fig. 4 and the range for (*b*)–(*d*) by eqn. (2) or the full lines of fig. 4. For the depth below the surface as abscissa the difference between (*a*) and the other curves would decrease at higher voltages and increase at lower voltages.

Stinchfield's case is weak, however, because Lenard's law is weak and because the dominating effects of scattering and straggling are not explicitly taken into account, though they are to some extent allowed for by the laws he uses. Equation (5), on the other hand, is based on a rather ruthless simplification of experimental material of moderate accuracy: still, it is probably the best 'law' available.

ACKNOWLEDGMENTS

The work described in this paper was carried out in the Physics Department of Birkbeck College, University of London. One of the authors (J. F.) was in receipt of a maintenance grant from the Department of Scientific and Industrial Research.

REFERENCES

BERGER, H., and PAUL, W., 1949, *Z. Phys.*, **126**, 422.

BETHE, A., 1930, *Ann. Phys., Lpz.*, **5**, 325.

BOHR, N., 1913, *Phil. Mag.*, **25**, 10; 1915, *Ibid.*, **30**, 581.

BOTHE, W., 1933, *Handb. Phys.*, XXII, **2** (2nd Edn.).

EDDY, C. E., 1928, *Proc. Camb. Phil. Soc.*, **25**, 50.

FRANKS, J., 1953, *Brit. J. Appl. Phys.*, **4**, 377.

GARLICK, G. F. J., 1949, *Luminescent Materials* (Oxford : Clarendon Press).

KOLLER, L. R., and ALDEN, E. D., 1951, *Phys. Rev.*, **83**, 684.

MADGWICK, E., 1927, *Proc. Camb. Phil. Soc.*, **23**, 970.

PALLUEL, P., 1947, *C. R. Acad. Sci., Paris*, **224**, 1492.

SCHONLAND, B. F. J., 1923, *Proc. Roy. Soc.* A, **104**, 235; 1925, *Ibid.*, **108**, 187.

STÖBER, F., 1924, *Z. Kristallogr.*, **61**, 299.

TAYLOR, C. J., REMLEY, M. E., JENTSCHKE, W. K., and KRUGER, P. G., 1951, *Phys. Rev.*, **83**, 170.

TERRILL, H. M., 1923, *Phys. Rev.*, **22**, 101; 1924, *Ibid.*, **24**, 616.

THOMSON, J. J., 1906, *Conduction of Electricity through Gases* (Cambridge : University Press), pp. 375 etc.

TSIEN, S. T., 1944, *Ann. Phys., Paris*, **19**, 327.

VARDER, R. W., 1915, *Phil. Mag.*, **29**, 725.

WHIDDINGTON, R., 1912, *Proc. Roy. Soc.* A, **86**, 360; 1914, *Ibid.*, **89**, 554.

WHITE, P., and MILLINGTON, G., 1928, *Proc. Roy. Soc.* A, **120**, 701.

WILLIAMS, E. J., 1931, *Proc. Roy. Soc.* A, **130**, 310.

Water Bells

By G. N. LANCE and R. L. PERRY
Department of Mathematics, King's College, London

Communicated by F. L. Hopwood; MS. received 31st August 1953

Abstract. The authors have studied theoretically the water bell phenomena, interest in which was revived recently by Hopwood. A numerical method of solution of the differential equation of the bell's surface is described and the results obtained are given in the form of graphs. It is shown that the shapes depend on three parameters and, in order to determine numerical magnitudes of these, experiments similar to Hopwood's were performed, during the course of which some photographs of the bells were obtained. Four of these are included. The paper ends with a section comparing the theoretical shapes with those obtained experimentally showing the agreement between the two to be good.

---

§ 1. INTRODUCTION

THE subject of water bells is by no means a modern one. The earliest reference to them is in papers by Savart (1833, 1834); a further study was made by Boussinesq (1869, 1913) and in later days the bells have been used to determine the surface tensions of various liquids. The subject was revived recently by Hopwood (1952) who described a new method of producing the bells and gave a description of the novel experimental results he obtained.

The present paper gives a theoretical treatment of the subject as well as describing some further experimental work. In the theoretical section it is shown that the shapes of the bells depend on three parameters; in order to get solutions of the equation involved it is necessary to know the numerical magnitude of these parameters. Hence it was decided to repeat the experiments in a slightly different way from that used by Hopwood.

§ 2. DESCRIPTION OF THE APPARATUS AND OF THE EXPERIMENTAL METHOD

The apparatus, shown in fig. 1, consists of a hollow brass tube to the top of which is fitted a cap in such a way that there is a slot between the cap and the top of the tube. The width of the slot can be varied from zero to a maximum value of 0·1 in. determined by the size of the tube.

There is a perforated disc about $1\frac{1}{2}$ in. from the top cap, the idea of which is to make the water project horizontally and steadily from the tube. The tube whose diameter is 1 in. is placed vertically in a trough of water so that the slot is about 2 in. above the water surface.

The experiment is performed in the following way. Water is passed through the tube, and if the bell so formed is punctured with the finger—thus equating air pressures inside and out—the shape shown in plate I is obtained. In the notation introduced below this is equivalent to putting $\alpha = 0$. The water flow and the size of the slot are now left unchanged but air bubbles are blown into the bell, without piercing the shell, thus making α negative. The bell changes shape gradually as the air is blown in—passing through a shape shown in plate II—until a very unstable shape is reached. It was impossible to photograph this state but as reported by Hopwood the water seems to rise, to the form of a cusp, above the level of the slot before collapsing into the initial stable shape with $\alpha = 0$, the air having leaked out at this apparent cusp. If air is withdrawn from the stable bell α becomes positive and a shape shown in plate III is formed. All the air can be withdrawn without any unstable shape being formed, the bell just travelling up the tube.

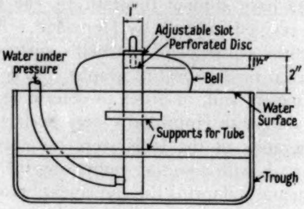


Fig. 1. Diagram of the apparatus.

Similar shapes can be formed by increasing and decreasing the water flow instead of blowing air into the bell; these were Hopwood's methods. However, reducing the water velocity is equivalent to blowing air in and vice versa: furthermore the method we employed is more convenient since it keeps two of the parameters fixed. The size of the slot is also important in that, for a given quantity of water flowing per second, if the slot size is decreased, the velocity of flow is increased and vice versa. Plate IV shows an interesting shape which is very much like a true bell, a slight bump appears near the water surface; this shape is only obtained when the size of the slot is quite large.

§ 3. Derivation of the Differential Equation of the Surface

The equations of motion were derived by Boussinesq (1869) but for convenience we give an outline of their derivation here. Figure 2 shows the coordinate system used.

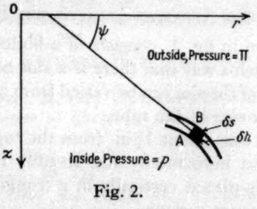


Fig. 2.

The continuity equation arises from the statement that the quantity of water Q flowing in unit time across any section AB perpendicular to the curve, is constant; thus

$$Q = 2\pi r h v \rho, \qquad \ldots\ldots(1)$$

where ρ is the density and h and v are the thickness of the shell and the velocity of the water at the section AB, respectively.

Consider the element of volume shown in fig. 2. Since there is no resultant surface tension force in a tangential direction, the tangential equation of motion is

$$h\delta s\,.\,r\delta\theta\,.\,\rho g\sin\psi = hr\delta s\,.\,\delta\theta\rho v\,dv/ds$$

or

$$g\sin\psi = v\,dv/ds. \qquad \ldots\ldots(2)$$

The equation of motion in the normal direction can be written as

$$\left(\frac{2T}{R}+\frac{2T\sin\psi}{r}+\Pi-p\right)+\rho hg\cos\psi=\frac{v^2}{R}\rho h, \qquad \ldots\ldots(3)$$

where T is the surface tension, Π and p are the air pressures outside and inside the bell, respectively, and R is the radius of curvature at AB in the vertical plane.

Now (2) may be written as $gdz = v\,dv$ and so by integration

$$2gz = v^2 - {v_0}^2, \qquad \ldots\ldots(4)$$

where v_0 is the initial velocity of projection. Writing $\alpha = 2\pi(\Pi-p)/Qv_0$ and $\beta = 4\pi T/Qv_0$ and using (1) we can write (3) as

$$v_0\beta\left(r\frac{d\psi}{ds}+\sin\psi\right)+\alpha rv_0+\frac{g\cos\psi}{v}=v\frac{d\psi}{ds}. \qquad \ldots\ldots(5)$$

Eliminating g from (4) and (5) we can obtain the equation used by Boussinesq (1869) in his derivation of the value of the surface tension

$$T=\frac{Q}{4\pi}\left(\frac{-\Delta(v\cos\psi)}{\Delta s-\Delta(r\cos\psi)}\right),$$

where $\Delta = d/d\psi$. However, to derive the differential equation of the surface we must eliminate v from (4) and (5) leading to

$$v_0\beta\left(\frac{z''r}{(1+z'^2)^{3/2}}+\frac{z'}{(1+z'^2)^{1/2}}\right)+\alpha rv_0+\frac{g}{(1+z'^2)^{1/2}(2gz+{v_0}^2)^{1/2}}=\frac{z''(2gz+{v_0}^2)^{1/2}}{(1+z'^2)^{3/2}}.$$

If we write $\gamma = g/{v_0}^2$ and $e^2 = 1 + 2\gamma z$ this may be more conveniently written as

$$1+z'^2\,\{(\gamma/e)+\beta z'+\alpha r\,(1+2'^2)^{1/2}\}=z''\,(e-\beta r)$$

or

$$R\left(\beta\sin\psi+\alpha r+\frac{\gamma\cos\psi}{e}\right)=e-\beta r. \qquad \ldots\ldots(6)$$

Several theoretical methods have been tried to solve this equation but have failed due to its complexity. Even if we make the further assumption that $\alpha = 0$, the equation still appears insoluble. We are therefore forced to seek a numerical method of solution.

§ 4. NUMERICAL SOLUTION OF THE DIFFERENTIAL EQUATION OF THE SURFACE

The equation of the surface of the bell is determined by a second order non-linear differential equation, (6), so it is unlikely that a complete solution, other than a numerical one, will ever be deduced. There is however one fact which makes a numerical solution particularly simple.

From (6) we see that an expression for the radius of curvature R is given directly; with this in mind we study the numerical solution of an equation of the form

$$R = f(r, z, z'). \qquad \ldots\ldots(7)$$

The method described is an extension of Euler's polygon method for first order differential equations. In fig. 3 the coordinates of the point P_0 namely (r_0, z_0)

are assumed known, the slope ψ_0 of the curve at P_0 is also assumed known. Our task is to determine the coordinates of a neighbouring point P_1 (r_1, z_1) and also the slope ψ_1 at P_1.

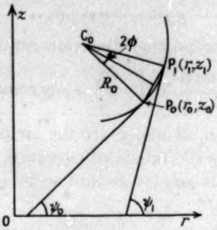


Fig. 3. Notation used in the numerical integration.

From (7) the radius of curvature at P_0 is calculated $R_0 = f(r_0, z_0, \tan\psi_0)$. This locates the centre of curvature C_0 although we do not need to find explicit expressions for its coordinates. Simple trigonometry shows that if we travel clockwise round the arc of a circle, centre C_0 and radius R_0 through an angular distance 2ϕ we arrive at a point whose coordinates are $r_0 + 2R_0 \sin\phi \cos(\psi_0 + \phi)$ and $z_0 + 2R_0 \sin\phi \sin(\psi_0 + \phi)$; the slope at this point is $\psi_0 + 2\phi$. The sign convention is important; in this instance, we say that ϕ is positive for clockwise rotation about C_0. Provided we keep ϕ small—1 or 2 degrees—we arrive at the point P_1 which is another point on the curve satisfying the differential equation (7). Thus we have shown that

$$\left.\begin{aligned} r_1 &= r_0 + 2R_0 \sin\phi \cos(\psi_0 + \phi), \\ z_1 &= z_0 + 2R_0 \sin\phi \sin(\psi_0 + \phi) \\ \text{and} \quad \psi_1 &= \psi_0 + 2\phi. \end{aligned}\right\} \qquad \ldots\ldots(8)$$

In general

$$\left.\begin{aligned} r_{n+1} &= r_n + 2R_n \sin\phi \cos(\psi_n + \phi), \\ z_{n+1} &= z_n + 2R_n \sin\phi \sin(\psi_n + \phi), \\ \psi_{n+1} &= \psi_n + 2\phi. \end{aligned}\right\} \qquad \ldots\ldots(9)$$

In the above equations the angle ϕ need not be kept constant: in the numerical work it was found convenient to vary it in such a way that the arc length P_nP_{n+1} remained about 0·2 cm. One further point needs emphasis; if at a point the radius of curvature is negative and if we are to proceed along the curve in the same direction as before we must travel clockwise round the point C_0; hence, according to our convention, ϕ must have a negative sign. The formulae (9) are true for both positive and negative R, ϕ and ψ. The errors involved in the above method are a little difficult to determine but we believe that they are smaller than those arising in Euler's method.

The general theory described above was applied to the particular equation (6) derived above

$$R = \frac{e - \beta r}{\beta \sin\psi + \alpha r + (\gamma \cos\psi)/e}.$$

Typical values of β and γ are obtained by assuming the following values: $T = 72\,\text{dyn cm}^{-1}$, $Q = 5$ litres min^{-1}, width of slot $= 0{\cdot}14$ cm. Thus the velocity of projection is $v = 90{\cdot}22\,\text{cm sec}^{-1}$, also $\beta = 0{\cdot}1203$ and $\gamma = 0{\cdot}1205$. Furthermore if we assume that the pressure difference across the shell cannot exceed 0·1 cm of water then $|\alpha| < 0{\cdot}08$.*

We choose the origin of coordinates such that at $r = 1$, $z = 0$ the water is projected horizontally i.e. $\psi = 0$. With these initial conditions five curves were obtained using the above values of β and γ in each case but α was taken to be $+0{\cdot}08$, 0, $-0{\cdot}03$, $-0{\cdot}04$, and $-0{\cdot}08$. The numerical integration was performed until the z coordinate reached $+3$, except in the case $\alpha = +0{\cdot}08$ when the integration was continued until the bell reached the tube again.

§ 5. COMPARISON OF EXPERIMENTAL AND THEORETICAL RESULTS

The theoretical curves are shown in fig. 4, the shapes shown, for $\alpha < 0$, are clearly not attainable in physical reality but we suggest that the water takes the path shown by dotted lines. The air escapes at the vertex or cusp and the bell collapses into the stable $\alpha = 0$ form.

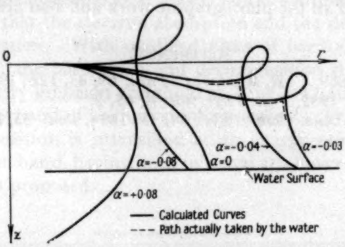


Fig. 4. Graphs of calculated bell shapes. Cases for which $\beta = 0{\cdot}1203$ and $\gamma = 0{\cdot}1205$ and α varying.

Unfortunately the authors did not have access to a 'Chattock tube' or similar apparatus for measuring small differences of pressure so were unable to determine α to any greater degree of accuracy than did Hopwood. However we suggest that probably the plates I, II and III correspond to the cases $\alpha = 0$, $-0{\cdot}04$ and $+0{\cdot}08$, respectively, shown in fig. 4. In order to obtain, theoretically, a shape such as plate IV v_0 would have to be reduced considerably, thus altering β and γ as well as α.

§ 6. CONCLUSIONS AND FUTURE DEVELOPMENTS

We have shown above how the differential equation of the water bell's surface can be integrated numerically for particular values of the parameters and how shapes corresponding very closely to the experimental ones are thus obtained. The loops which are derived theoretically are rejected on the grounds that in practice the water cannot cross itself—the curves shown dotted would actually appear. With this assumption and bearing in mind the fact that no measurements of pressure could be made the similarity is quite encouraging.

Several interesting questions now present themselves and they will be left as the subject for a future paper. Firstly, how does the angle of projection affect the bell shape? Theoretically the problem is now to solve the equation as before but, at the point $r=1$, $z=0$, instead of $\psi=0$ we must take ψ equal to, say, 45°. The apparatus can be modified by turning the inside corners off the cap and the side of the tube.

---

* In a private communication to the authors Professor F. L. Hopwood has stated that an error occurred in his paper. It was asserted that the pressure difference was less than 0·1 mm of water; this should have read less than 0·1 cm of water.

Secondly, the bell shapes considered above may be called steady since time does not enter the problem. It should be possible to make the cap a sliding fit on the top of the tube and allow the cap to oscillate up and down as the water flows; we believe that this would produce small waves on the bell surface which would considerably enhance their beauty! The unsteady problem is however rather more difficult to solve theoretically since the differential equation of the surface would be altered and would contain time as an additional independent variable.

ACKNOWLEDGMENTS

The authors wish to thank Professor F. L. Hopwood for the helpful comments he has made during the preparation of this paper. Our thanks are also due to Mr. I. Jordan who did all the photographic work and also prepared the diagrams.

REFERENCES

BOUSSINESQ, J., 1869, *C. R. Acad. Sci., Paris,* **69,** 45, 128; 1913, *Ibid.,* **157,** 89.
HOPWOOD, F. L., 1952, *Proc. Phys. Soc.* B, **65,** 2.
SAVART, F., 1833, *Ann. Chim. Phys.,* **54,** 55; 1834, *Ibid.,* **55,** 257.

The Decomposition of Thin Films on Bombardment with Slow Electrons

By D. A. WRIGHT and J. WOODS
Research Laboratories, General Electric Company, Wembley, Middlesex

Communicated by the Director, G.E.C. Research Laboratories; MS. received 3rd March 1953, *and in amended form 26th June* 1953

Abstract. The thresholds are determined for the incidence of electron absorption, decomposition and secondary emission for several types of thin film bombarded with slow electrons. With crystalline barium oxide, electron absorption and decomposition both set in at a bombarding energy near 4 ev, and are intensified at 7 ev. Secondary emission is initiated at 5 ev and increases at 7 ev and 10 ev. The optical absorption threshold is at 3·8 ev, and is attributed to exciton formation; it seems likely that the electron absorption and the decomposition are also due to exciton formation. With oxidized films of barium, crystalline barium oxide is not formed; the absorption and decomposition thresholds are at 7 ev, and the main secondary emission threshold at 10 ev. Decomposition of barium chloride sets in at 5 ev, and of barium sulphate at 7 ev. With films of the alkali halides, the decomposition is intensified at an energy corresponding with the first optical absorption band, having been initiated at a lower energy. A tentative explanation of this is proposed.

---

§ 1. Introduction

SEVERAL writers (Wagner 1931, Headrick and Lederer 1936, Hamaker, Bruining and Aten 1947, Dart 1950, König 1951) have discussed the decomposition of thin films of poorly conducting compounds under electron bombardment. Electron energy thresholds have been encountered, at which the decomposition is first detected. Threshold values from the literature are as follows (Jacobs 1946, Jacobs and Dobischek 1951): FeO 5·5, CuO 2·0, NiO 2·5, WO_2 4·5, MoO_3 7·4, Ta_2O_5 4·6, KCl 4·5, ZrO_2 1·7 ev.

It was pointed out by Jacobs (1946) that with some of these compounds the decomposition threshold occurs at an electron energy equivalent to the heat of formation of the compound. Assuming this to be general, the threshold energies observed in certain pentodes operated as diodes (Metson 1949) could be attributed to specific barium compounds.

It is important to know whether the correlation of threshold energy with heat of formation is reliable, both because of its bearing on solid state physics, and because if so, observation of threshold energies could be undertaken either as a means of analysing films of unknown composition or of determining the heat of formation of films of known composition.

The decomposition of barium oxide under electron bombardment was found by the writers to be important as affecting secondary emission (Woods and Wright 1952, 1953). The present work started therefore as a study of the threshold energies of decomposition and of secondary emission for barium oxide.

It was extended to include other types of thin films when it became evident that with barium oxide the decomposition threshold did not in fact correspond with the heat of formation. The disagreement was greater than 1 ev, and was found to be even larger with barium chloride and barium sulphate.

The experiments have been carried out using diode or triode systems, with an oxide coated cathode at low temperature (400°c) as the source of the bombarding electrons. The effects of the electrons on the film have been studied by observing the variations in cathode emission and in the valve characteristics. When the film is dissociated, the products modify the cathode emission. Various secondary effects have been met, which needed study in order to clarify the effects of bombardment. The secondary phenomena include electron reflection and charging of the surface of the film.

The experimental procedure and results are presented in some detail for the films of barium oxide, as this was the compound of particular interest. The results with other types of film are described more briefly. This part of the work was undertaken in order to find an alternative explanation of the decomposition threshold, when it had been found that this did not correspond with the heat of formation. Several oxides and alkali halides were studied. The experimental procedure was as with the barium oxide films.

§ 2. EXPERIMENTAL PROCEDURE

The valves used in this investigation were simple diodes and triodes with special anode systems. Two plane nickel anodes were mounted on a cylindrical framework which could be rotated about its axis when a magnet was brought near to an arm of soft iron attached to the framework. The anodes were diametrically opposite to one another, and either could be made to face the cathode. Four-watt box cathodes of O-nickel sprayed with barium–strontium carbonate were used in all the valves. (O-nickel is nickel to the following specification: Ni + Co 99·5% min, Mg 0·07–0·15%, Co $<$0·5%, Mn $<$0·15%, Fe $<$0·2%, Si $<$0·1%, Cu $<$0·1%, S $<$0·005%, C $<$0·04%.) In the triodes the grids, which were planar, carried ten 0·1 mm molybdenum wires per centimetre. This small number was to allow a large proportion of the cathode current to reach the anode when a valve was operated with the grid at a higher positive potential than the anode. When barium oxide was to be studied, the valves carried a helix of platinum wire sprayed with barium carbonate, from which, after conversion to the oxide, barium oxide was evaporated on to one anode. The halides and the barium sulphate were also evaporated from a platinum helix. Other metal oxides were prepared by oxidation of the appropriate anode metal, molybdenum replacing nickel when molybdenum oxide was to be studied.

The valves were pumped on a mercury-in-glass diffusion pump with liquid air trap. After baking at 400°c for half an hour, the getter and metal electrodes were outgassed by eddy current heating. The cathode coating was converted from carbonate to oxide, as was the barium carbonate on the platinum spiral. Following further eddy current heating of all metal parts, the grids in the triodes were heated to 800°c by electron bombardment. The valves were then sealed off and gettered. The cathodes were activated by maintaining them at a temperature of 900°c and taking a space charge limited current of 120 mA cm^{-2}. For this operation, which lasted twenty minutes, the triode grids were strapped to the anodes. During cathode outgassing and activation, and at all times when the cathode was at a high temperature, the same anode always faced it. The second anode and the platinum spiral were then shielded from any material evaporating from the cathode. After the activation a thin film of barium oxide was evaporated on to the clean anode which was then rotated to face the cathode. Experiments could then be begun.

With the diodes the cathode was operated at about 400°c. First a retarding field run was made to determine the contact potential difference between anode and cathode. After this, the anode voltage V_a was increased from zero in 1 volt steps, and at each setting the anode current I_a was noted during a three minute waiting period.

With the triodes the cathode was also operated at 400°c. The grid potential V_g was kept constant at 75 volts while the anode potential V_a was varied from 0 to 30 volts. With this arrangement the grid accelerated electrons from the cathode; of these some were intercepted, but the rest penetrated the grid and bombarded the anode. Secondary electrons and reflected primaries arising at the anode as a result of this bombardment were collected by the grid. At most settings of V_a the grid and anode currents varied with time, so that (I_g, V_a) and (I_a, V_a) curves had to be determined rapidly. In some valves the time changes were so large that these characteristics could not be found. A second run was always made, during which I_a and I_g were noted during a three minute wait at each setting of V_a. The cathode emission was checked before and after each such interval by earthing the anode, swinging it away from the cathode, and measuring the grid current.

The cathodes in these valves were run at low temperatures for three reasons: (i) to ensure that no material evaporated on to the barium oxide film under test, (ii) to reduce space charges to a minimum, and (iii) to maintain a high sensitivity of the cathode to any poisoning agency released from the anode during the electron bombardment. The thermionic emission at the operating temperature was between 10 and $100\,\mu\text{A}\,\text{cm}^{-2}$, a current easily saturated with a grid potential of 75 v. The secondary currents from anode to grid were also completely saturated.

§ 3. CONTACT POTENTIAL DIFFERENCE AND ELECTRON REFLECTION

In order to estimate the energies of electrons arriving at the anode of either type of valve it is necessary to know the contact potential difference V_0 between anode and cathode. It is convenient to discuss first the way in which V_0 was found for the diode valves.

In the absence of space charge and when the true anode potential $V = V_a - V_0$ is less than zero (the retarding field region), I_a should depend on V according to the relation $I_a = I_s \exp(\boldsymbol{e}V/\boldsymbol{k}T)$, where I_s is the saturated emission, $\boldsymbol{e}$ is the charge on an electron, $\boldsymbol{k}$ is Boltzmann's constant and T is the cathode temperature.

At $V = 0$, $I_a = I_s$, and for higher anode potentials I_a increases slowly, according to the Schottky law $\log I_a \propto V^{1/2}$. In practice, with metal anodes and low cathode temperatures, two straight lines are usually obtained when $\log I_a$ is plotted against V_a. These lines intersect close to the true zero of anode potential. The error introduced by taking the point of intersection as the true zero is not large, being of the order of 0·1 ev. It arises because $\log I_a$ is plotted against V_a instead of against $V^{1/2}$ in the Schottky region.

Figure 1 shows $(\log I_a, V_a)$ curves for one of the rotating anode diodes with its cathode at 400°c. Curve A was taken with the nickel anode facing the cathode, and curve B with the other anode in position carrying a film of BaO 5×10^{-6} cm thick. The thicknesses of the various BaO films were estimated from the interference colours formed by reflection. With the nickel anode in position two straight lines were obtained when $\log I_a$ was plotted against V_a as curve A shows. The abscissa of the point of intersection of these lines denotes the true zero of anode potential, while the ordinate gives a measure of the saturated emission I_s. With a film of BaO covering the anode the $(\log I_a, V_a)$ plot did not yield two straight lines as curve B, fig. 1, demonstrates. At very low currents, curve B becomes parallel to the proper retarding field line. By measuring the displacement between these two parallel lines and knowing the contact potential difference between the cathode and the nickel anode it was possible to calculate the contact potential difference between the cathode and the BaO film. This method was used by Moore and Allison (1950) when working with an anode covered with an evaporated layer of strontium oxide.

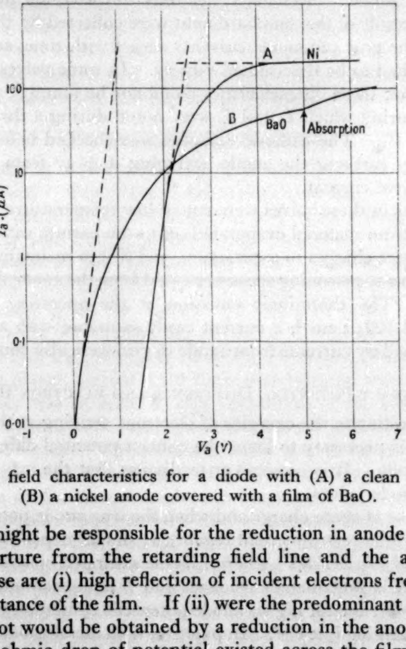


Fig. 1. Retarding field characteristics for a diode with (A) a clean nickel anode and (B) a nickel anode covered with a film of BaO.

Two effects might be responsible for the reduction in anode current and the consequent departure from the retarding field line and the apparent lack of saturation. These are (i) high reflection of incident electrons from the BaO film and (ii) high resistance of the film. If (ii) were the predominant feature a normal retarding field plot would be obtained by a reduction in the anode current, such that a negligible ohmic drop of potential existed across the film. In fact when the current was reduced 200 times, there was still the same deviation from the retarding field line and the same lack of saturation. The experiments on the triodes also support the view that, with the currents used here, resistance effects of films thinner than 10^{-5} cm are negligible. It is concluded, therefore, that the departure of curve B from the theoretical curve is due to the reflection of a considerable number of low energy electrons incident on the oxide.

Curve B, fig. 1, shows an upward kink at the point where the electrons have an energy of 4 ev. This happens at $V_a = 5$ v in curve B where the correction for contact potential difference is 1 ev. The kink at 4 ev is due to the onset of absorption of incident electrons when their energy reaches this value. Experiments on the triode valves to be described below confirm the existence of this absorption threshold.

§ 4. SECONDARY EMISSION AND ELECTRON REFLECTION

In the triode valves with the grid operated at a higher potential than the anode, the grid current is composed of two parts, that function of the cathode emission which is intercepted by the grid wires, and the secondary current from the anode. Curve A of fig. 2 shows the grid current as a function of V_a for a film of BaO 2×10^{-6} cm thick. Curve B is the corresponding anode current plot. These curves are uncorrected for contact potential difference. All the curves obtained with films thinner than 10^{-5} cm show essentially similar characteristics. With no potential applied between anode and cathode all the measurable current flows to the grid. As V_a is made positive, current flows to the anode and I_g falls. When the electrons reaching the anode possess an energy of 4 ev a large drop in I_g occurs, accompanied by a corresponding increase in I_a. This happens at $V_a = 5$ v in fig. 2, where the correction for contact potential difference is 1 v.

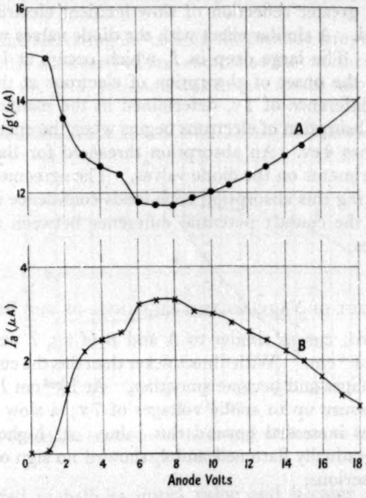


Fig. 2. Triode characteristics at low anode voltage. Grid at 75 v.

At an energy of 5 ev (6 v in fig. 2), the rapid fall in I_g is arrested, and the curve becomes nearly horizontal. Thus either the reflection coefficient increases again, or more probably there is an initiation of secondary emission, at an energy of 5 ev. There is a further kink at 7 ev energy, where I_g begins to increase, and at 10 ev a kink is frequently seen where the slope increases further, though this is not shown in fig. 2. Thus the secondary emission, which is probably initiated at 5 ev bombarding energy, increases discontinuously at 7 ev and at 10 ev. The curves were rarely continued beyond $V_a = 15$ v because considerable decomposition of the oxide occurred at these anode potentials and the cathode was rapidly poisoned.

With these valves the contact potential difference between cathode and anode was allowed for by taking as the reference point from which to measure electron energies, that value of V_a at which the (I_g, V_a) curve ceases to be horizontal and begins its downward trend. This is seen to happen at $V_a = 1$ v in fig. 2. The reason for this procedure is as follows. At very low anode potentials where no measurable anode current flows, and I_g is sensibly constant, any electrons which penetrate the grid enter a strong retarding field and return to the grid. When V_a is increased so that the first measurable anode current flows, some of the fastest electrons overcome the retarding field and reach the anode. Since the spread in the initial velocities of electrons emitted by the cathode is small this point may be taken as the true zero, and energies computed for higher anode potentials using this zero will represent the energies of the fastest electrons reaching the anode. The bulk of the electrons will have slightly lower energies, but the maximum error is less than 0·5 ev.

As the true potential of the anode is increased from zero I_g begins to fall, but it does not fall so rapidly as when an anode of clean nickel is used. This is because there is a greater reflection of slow incident electrons from BaO than there is from nickel. A similar effect with the diode valves was described in the preceding section. The large drop in I_g which occurs at $V_a = 5$ v in curve A of fig. 2 is due to the onset of absorption of electrons at this point. Using a contact potential difference of 1 v, determined in the manner described above, it is seen that this absorption of electrons begins when the energy of the bombarding electrons reaches 4 ev. An absorption threshold for BaO at 4 ev was also found during experiments on the diode valves. The agreement between the two methods in predicting this absorption edge lends confidence in the method used here to determine the contact potential difference between anode and cathode in the triode valves.

§ 5. EFFECT OF VARYING THE THICKNESS OF THE BaO FILM

As already stated, curves similar to A and B of fig. 2 were obtained for all films thinner than 10^{-5} cm. With films thicker than this the curves were displaced towards higher voltages and became smoother. At 10^{-4} cm I_g was constant and I_a too small to measure up to anode voltages of 7 v; a slow decrease in I_a was observed as V_a was increased beyond this value. At higher anode potentials the (I_g, V_a) curve gradually flattened and I_g showed no sign of increasing before poisoning became serious.

The reason for this excursion to higher voltages and the accompanying flattening with thicker films is that the resistance of the barium oxide becomes large enough to play a significant part. When this happens the potential at the oxide surface is much lower than that of the underlying metal anode because of the ohmic drop in potential across the film. As a result the (I_g, V_a) curves are displaced to higher voltages and are effectively spread out over a wider voltage range, causing the observed smoothness. That this is the correct explanation was established in the following way. It is known that the secondary emission coefficient δ of BaO is unity when the bombarding electrons have energies of 30 ev (Woods and Wright 1952). With lower energies $\delta < 1$, and with higher energies $\delta > 1$. When a film with sufficient resistance to support a surface charge is bombarded with 25 ev electrons $\delta < 1$ and the film acquires a negative surface charge. When V_a is increased rapidly to find the point at which $\delta = 1$, an over-estimate of the true value of V_a is obtained, due to the inability of the negative charge to leak away fast enough. By adopting the opposite procedure and applying a beam of 35 ev electrons to an uncharged film, so that $\delta > 1$, the film charges positively. When V_a is reduced rapidly to the point at which $\delta = 1$, an under-estimate of the proper value of the required anode potential is obtained. If, therefore, these two ways of determining the value of V_a at which $\delta = 1$ do not agree, it is clear that charging effects, due to the high resistance of the film, are playing an important part. Using this technique no measurable effect (i.e. as large as 0·2 ev) due to high resistance was found for films thinner than 10^{-5} cm, while thicker films showed an increasing tendency to charge as the thickness was increased.

§ 6. THERMIONIC EMISSION OF THE OXIDE CATHODE

Accurate (I_a, V_a) and (I_g, V_a) curves can only be obtained if the thermionic emission of the cathode remains constant. In fact, changes in cathode current did take place, but in most valves the changes were small provided V_a did not exceed 15 v.

After pumping and activating the triodes, the cathode temperature was reduced to 400°C. The anodes were earthed and rotated so that neither faced the cathode. With 75 v applied to the grid, emission was drawn for 20 minutes. In this time the grid current fell to half its initial value and became steady. With the cathode in this state the BaO film was rotated to face the cathode and (I_a, V_a) and (I_g, V_a) characteristics were taken. After this the emission to the grid with the anode earthed was measured, then V_a was increased from zero in 1 volt steps, with a waiting period of 3 minutes at each setting of V_a. The thermionic emission was checked before and after each 3 minute period. The curves in fig. 3 show the way in which the emission depends on the anode voltage for a number of valves.

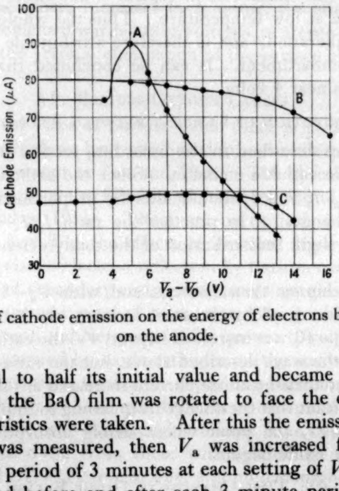


Fig. 3. Dependence of cathode emission on the energy of electrons bombarding a BaO film on the anode.

For $V_a - V_0$ less than 4 v there is no change in emission with time. When the bombarding electrons have an energy of 4 ev the emission alters, and from fig. 3 it is clear that the emission may either rise or fall. If the first effect is a decrease, further increases in anode potential enhance the effect. On the other hand, if an increase occurs first, increasing V_a may produce further increases in emission. Ultimately, however, the increase is superseded by a decay. Once a rising emission has turned into a falling one in this way, higher anode potentials lead to further decays. The value of $V_a - V_0$ at which the rise goes over to a decay is of no particular significance, since this happens at $V_a - V_0 = 4$ v if the experiment is continued for long enough. The curve in fig. 4, which shows the variation in emission while bombarding with 6 ev electrons, demonstrates this point. In this run it took $2\frac{1}{2}$ hours for the rise to turn to a decay.

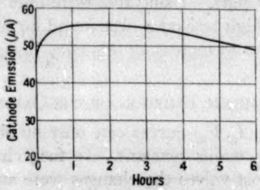


Fig. 4. Variation of emission with time while bombarding a BaO film at 6 ev energy.

The increase in emission which sometimes preceded the fall when 4 ev was reached was unexpected, and appeared at first to suggest that two processes were involved. However, the work of Liebold, quoted by Herrmann and Wagener (1951), showed that an initial rise is typical of the effect of a low pressure of oxygen on a cathode at low temperature. Thus the whole of this effect can be caused by oxygen liberation, and is therefore attributable to decomposition of the oxide under bombardment. It can be concluded that the decomposition threshold for BaO is near 4 volts.

§ 7. DISTRIBUTION OF CURRENT BETWEEN ANODE AND GRID

The time changes described in the preceding section were solely concerned with the variations in cathode emission. Since the cathode current is equal to the sum of I_a and I_g, I_a and I_g should alter by proportional amounts when the cathode emission changes. In practice the ratio I_g/I_a often alters slightly, which means that a slight redistribution of the total current between anode and grid takes place.

With the films thinner than 10^{-5} cm, and with $V_a - V_0 < 4$ v, I_a often fell with time while I_g increased by the same amount, the total current remaining constant. When $V_a - V_0$ was increased beyond 4 v, the cathode current changed with time in one of the ways described above, but the ratio I_g/I_a decreased with time. This indicates that the anode current increased at the expense of the grid current. It is significant that the change from falling to improving anode current sets in at $V_a - V_0 = 4$ v, the point where strong absorption of electrons and decomposition of the oxide begins.

§ 8. TIME CHANGES IN DIODE VALVES

The time changes observed in the diodes correspond exactly with those found with the triodes. With incident electrons of 4 ev energy the barium oxide is decomposed. The decomposition products react with the cathode and either increasing or decreasing emission is observed. In these valves the emission was nearly always poisoned. This was perhaps because the experiments were performed immediately after activating the cathode, whereas with the triodes emission was drawn to the grid first, and a large decay ensued. In consequence the state of the cathodes was different in the two types of valves.

In the diodes time changes in I_a not associated with any variation in cathode emission were also found. At low anode voltages and with the clean nickel anode in position, I_a decreased very slowly with time. With the BaO film facing the cathode and with the same anode voltage, the anode current was much lower, due to reflection, and decayed much more rapidly. When V_a was increased an anode current rising with time was observed provided $V_a - V_0 > 4$ v. This increase was superimposed on any changes in I_a due to changes in cathode emission, thus in some valves I_a actually increased while the cathode emission fell. In order that I_a should increase with time in this manner, with true anode voltages greater than 4, it is necessary for a decay to have taken place at a lower anode potential. If a voltage greater than 4 was applied to a valve left standing idle for some time, a decay was seen. An increase in I_a with time was observed as soon as V_a was increased. Under no circumstances did I_a increase with time with $V_a - V_0$ less than 4 v.

When a retarding field curve was taken for a diode with its anode covered with BaO, a curve such as B, fig. 1, was obtained. When the retarding field measurements were repeated, after setting the true anode potential at 3 v and allowing a decay in I_a to proceed, the curve was found to have been displaced to higher voltages by a few tenths of a volt. This indicates that the anode became more electronegative, probably by acquiring a surface layer of negative charge. With higher anode voltages, where increases in I_a occur, the effect on the retarding field curve is a displacement towards lower voltages, nearer to its original position. This corresponds to the gradual elimination of the surface charge.

§ 9. BARIUM–OXYGEN LAYERS

With the evaporated films of barium oxide there is an apparent increase in reflection coefficient, which we have attributed to a secondary emission threshold, at 5 ev, and an enhancement at 7 ev and 10 ev. The threshold at 5 ev is to be expected since the depth of the filled band has been found optically to be 5 ev (Apker, Taft and Dickey 1951). There are no optical absorption figures for energies higher than 5 ev, but the effects at 7 ev and 10 ev are consistent with Rudberg's work on inelastic electron reflection by barium oxide (Rudberg 1936).

The development of secondary emission at an energy greater than the optical absorption threshold seems typical of many materials (Jonker 1947). However both Jonker and Matheson and Nergaard (1951) quote a 10 volt secondary emission threshold for films deposited by oxide cathodes on the surrounding electrodes.

We have investigated films formed by depositing barium on the nickel anode, and then oxidizing, usually by heating at 400°c in a pressure of 10^{-2} mm of oxygen, and have compared these with the deposits formed from oxide cathodes. The results are similar and differ from those with evaporated BaO. There is no change in slope in the (I_g, V_a) curve at 4 ev, and no decomposition effects can be detected at this energy. The plot becomes horizontal at 5 ev, and then at 7 ev both decomposition and absorption are pronounced. The value of I_g falls between 7 and 10 ev, and then rises sharply, so that 10 ev is the threshold for the appearance of secondary emission. This agrees with Jonker and with Matheson and Nergaard. Some studies of barium oxide evaporated on to different substrates have shown decomposition initiated at 4 ev, and intensified at 7 ev, so that it seems clear that the question of which features of the (I_g, V_a) curves are dominant, and where decomposition is most marked, vary with the physical form of the oxide. Whereas barium oxide evaporated from platinum gives an electron diffraction pattern of $Ba\,CO_3$ after opening to air and introducing into the diffraction camera, the films formed by oxidation of barium show no identifiable pattern (other than that of nickel, which is sometimes detected). Thus these films seem to consist of an amorphous complex of oxygen and barium, and not of crystalline BaO. The deposits from oxide cathodes are presumably similar, having similar properties on bombardment. Any secondary emission increase which may occur from these films at 7 ev is masked by the absorption increase at the onset of decomposition. The fact that decomposition with a threshold at 7 ev occurs in both types of film, but with a threshold at 4 ev only in crystalline films, is an important result which will not however be pursued here. It should be noted that an evaporated film of barium, after oxidation, has a much lower thermionic emission than crystalline BaO formed by evaporation from platinum. These latter films, incidentally, were found to be highly orientated when examined by electron diffraction after a little heat treatment had been given following deposition. Heating to about 700°c destroyed the orientation, but had no effect on thermionic emission or on the decomposition and absorption threshold.

§ 10. FILMS OTHER THAN BaO

Measurement of absorption and decomposition thresholds have been made in the same way on thin films of various compounds.* Secondary emission thresholds have been measured only with barium compounds. The results of this work are shown in the table. It will be seen that with KCl, MoO_3 and NiO we find decomposition thresholds similar to those observed by Jacobs, and that with these and with NaCl also this threshold energy is similar to the heat of formation. With NiO and MoO_3 however this can have little significance as the electron affinity must be quite large. With NaCl and KCl, the electron affinity is small, and the similarity might be significant. We believe however that this is fortuitous, for the reasons given below.

	(1)	(2)	(3)		(4)
			(*a*)	(*b*)	
BaO	4·0	5·4	4·0	3·8	5·0, 7·0, 10·0
NaCl	4·5	4·25	7·7	7·8	
KCl	4·0	4·5	6·5	7·5	
KI	2·5	3·5	5·5	5·4	
NiO	2·5	2·5	—	—	
MoO_3	8·0	7·4	—	—	
$BaCl_2$	5	8·9	—	—	10·5
$BaSO_4$	7	15·0	—	—	12

(1) First threshold: energy for decomposition (ev); (2) heat of formation (ev); (3) main absorption threshold (ev): (*a*) electron, (*b*) optical; (4) energy giving increase in apparent reflection, i.e. of reflection plus secondary emission (ev).

With the alkali halides, the main electron absorption threshold occurs at an energy near the optical absorption peak, as observed by Hilsch (1932). The decomposition threshold occurs at an energy considerably lower than this main absorption threshold. With each halide however there is a small absorption effect at the onset of decomposition. A corresponding behaviour was observed by Jacobs and Dobischek (1951) with KCl. There is moreover with each halide an intensification of the decomposition near the main absorption threshold, as estimated by the rate of poisoning of the cathode emission.

---

* Comparable results on alkali halides were obtained by Jacobs, Martin and Brand (1952).

§ 11. THE DECOMPOSITION PROCESS

The decomposition process investigated in this work results in the electronegative constituent being liberated from the thin film. The oxygen or halogen so freed moves to the cathode and alters its emission. It is of interest to consider the mechanisms which might account for this gas liberation.

The first stage in decomposition is likely to be the production of an electronic transition either (*a*) forming an exciton, or (*b*) forming a free electron and a free hole, or (*c*) tranferring an electron directly from a negative to a positive ion. (*a*) could lead to decomposition either following thermal dissociation of the exciton, so that the process thereafter becomes equivalent to (*b*), or following an interaction between the exciton and lattice vacancies. A process of this latter kind has been found applicable to the photolysis of barium and potassium azide (Jacobs and Tompkins 1952). Following (*b*), the migration of a hole to the surface could form an oxygen or halogen atom at the surface, while (*c*) occurring at the surface would also form atoms there. In a highly ionic compound, the formation of atoms at the surface of the crystal could lead fairly readily to their escape, though this would be much less likely with a more covalent type of bond. Thus with the alkali halides formation of surface atoms would give rapid halogen formation, while with barium oxide the freeing of oxygen would occur less readily.

The results on barium oxide show that the threshold energy for decomposition is very close to that for electron absorption, while both are very close to the optical absorption threshold (Tyler 1949, Taft and Dickey 1950, Tyler and Sproull 1951, de Vore and Dewdney 1951). The first optical absorption band is attributed to exciton formation, so that process (*a*) may be applicable here, either followed by thermal dissociation, or as a result of direct interaction with vacancies near the surface. There is however a possible alternative explanation, as suggested below.

With the alkali halides, the first main optical absorption band is also attributed to exciton formation, so that again process (*a*) may be involved, applying this time however to the enhancement of decomposition which occurs when the electron energy reaches the absorption band energy. The first decomposition threshold, occurring in the alkali halides at an energy below that at which optical absorption occurs in the pure crystal, is more difficult to explain. It is of course possible that any of the processes (*a*), (*b*) or (*c*) occur at the surface with an energy less than that involved within the crystal lattice, and that the slow electrons are causing decomposition only in the surface layer. The coincidence of the electron absorption threshold with the optical absorption band makes it unlikely however that this energy difference is appreciable. The energy for process (*c*) is not likely to be less than that for (*a*), if transfer at normal lattice sites is considered.

Transitions could however occur in an imperfect crystal, involving energy less than that required to form excitons, and capable of initiating decomposition. The first of these is the direct transfer of an electron from the valence band of the chlorine ions to a negative ion vacancy, forming an F-centre and a free hole.

Then this hole migrating to the surface could free a halogen atom. The converse of this is the transition which frees an electron from a chlorine ion in the immediate neighbourhood of a positive ion vacancy, forming a trapped hole at the vacancy, i.e. a V-centre, and a free electron. The V-centre would lead to decomposition if it moved to the surface, or if it were formed there at a kink. Usually the F-centres and V-centres are visualized as formed following either exciton formation, or formation of an electron–hole pair, followed by trapping. The direct transition is however a possibility, and the first one, forming an F-centre, has been recognized recently in studies of alkali halides at low temperature. It produces an absorption band designated the α band (Dexter 1951, Delbecq, Pringsheim and Yuster 1951, 1952, Martienssen 1952, Martienssen and Pohl 1952). This has been studied most fully in KI and KBr, and occurs in the long wavelength tail of the first absorption band, i.e. the exciton forming band. It seems possible that the initiation of decomposition in BaO, and the intensification of decomposition in the alkali halides, both occurring at the main optical absorption band energy, do not coincide exactly with this band, and are not due to exciton formation (process (*a*) above) but to transitions of the α-band type, forming free holes.

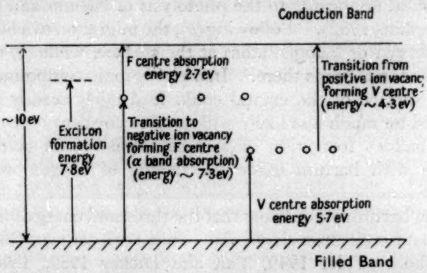

Conduction Band
~10 ev
Exciton formation energy 7·8eV
F centre absorption energy 2·7 eV
Transition to negative ion vacancy forming F centre (α band absorption) (energy ~ 7·3eV)
Transition from positive ion vacanc forming V centre (energy ~ 4·3 eV)
V centre absorption energy 5·7 eV
Filled Band

Fig. 5. Transitions in sodium chloride crystals.

The converse process forming V-centres does not appear to have been considered or recognized previously. It is suggested here very tentatively, on the grounds that the energy involved if such a process occurred would correspond with the observed decomposition threshold energy. The energy required to form a V-centre by this process would be obtained by subtracting from the width of the forbidden gap the energy corresponding with the absorption band observed in the alkali halide when it contains V-centres. In NaCl this is 5·7 ev, and the forbidden gap width is 10 ev, so that the energy of the postulated transitions would be 4·3 ev. The transitions are illustrated in fig. 5. Very similar figures apply for KCl. With KI, the forbidden gap width is 6·6 ev, and the V-centre absorption band occurs at 4 ev, so that the suggestion transition would occur at 2·6 ev. Thus in all three cases the energy corresponds with that at the decomposition threshold.

Indirect support for the occurrence of this type of transition comes also from the recent work on the electrolysis of NaCl (Hacskaylo and Groetzinger 1952). This work showed that electrolysis formed lattice vacancies, which were described as 'colour centre precursors', i.e. they were visualized in the same way as the vacancies we have postulated as present in the evaporated films. When these vacancies were present, there were absorption bands at 226 and 285 μ. The second of these had not previously been seen in NaCl, and the notable fact is that the corresponding energy is 4·34 ev. This we suggest is also the energy producing an electron transition from the Cl ion at a positive ion vacancy to the conduction band, forming a V-centre.

Thus is seems possible that the first stage in decomposition is a transition at or near the surface involving a lattice defect. The indirect support given for this suggestion is of course far from conclusive, but justifies giving it further attention in future work. It leads incidentally to the further speculation that other effects, such as the increase in photoelectric emission on irradiation (Dickey and Taft 1950, Apker and Taft 1951), might be due also to transitions of this type, as may the inflection at 5·3 ev in the long-wavelength tail of the photoelectric response of potassium iodide (Apker and Taft 1950).

§ 12. CONCLUSIONS

With barium oxide the threshold of decomposition is the same as for electron absorption, and occurs near 4 ev. This corresponds closely with the optical absorption threshold at 3·8 ev, attributed to exciton formation. At lower energy the reflection coefficient is large, and the surface takes a negative charge; above 4 ev some of this charge is destroyed. Secondary emission is initiated at 5 ev and increases abruptly at 7 ev and again at 10 ev. There is also an increase in absorption and decomposition at 7 ev.

The decomposition threshold for BaO does not occur at an energy similar to the heat of formation. This lack of correspondence is even more marked with $BaCl_2$ and with $BaSO_4$. An apparent correspondence for NiO and MoO_3 cannot be genuine, since the electron affinity must not be neglected.

The correspondence between heat of formation and decomposition threshold in NaCl and KCl is believed to be fortuitous, and does not occur with KI. The threshold should correspond with an electron transition which converts the halogen ion to an atom which is then able to escape since the binding is primarily ionic. Whereas with BaO the decomposition is initiated at an energy similar to that required for exciton formation, the corresponding energy produces with the alkali halides an enhancement of a decomposition process initiated at considerably lower energy.

This low energy decomposition may be due to electronic transitions occurring at or very near the surface, in the neighbourhood of lattice defects, freeing an electron from a halogen ion in the immediate vicinity of a positive ion vacancy. The higher energy decomposition may be due either to exciton formation, or to a process analogous to that postulated for the lower energy decomposition. This process would be the transfer of an electron from a halogen ion (or oxygen in BaO) to a neighbouring negative ion vacancy.

ACKNOWLEDGMENT

In conclusion the authors desire to tender their acknowledgments to the M.-O. Valve Co. Ltd., on whose behalf the work described in this publication was carried out.

REFERENCES

APKER, L., and TAFT, E., 1950, *Phys. Rev.*, **79**, 964; 1951, *Ibid.*, **81**, 698, **82**, 214.

APKER, L., TAFT, E., and DICKEY, J., 1951, *Phys. Rev.*, **84**, 508.

DART, F. E., 1950, *Phys. Rev.*, **78**, 761.

DELBECQ, C., PRINGSHEIM, P., and YUSTER, P., 1951, *J. Chem. Phys.*, **19**, 574; 1952, *Ibid.*, **20**, 746.

DEXTER, D. L., 1951, *Phys. Rev.*, **83**, 1044.

DICKEY, J. E., and TAFT, E. A., 1950, *Phys. Rev.*, **80**, 308.

HACSKAYLO, H., and GROETZINGER, G., 1952, *Phys. Rev.*, **87**, 789.

HAMAKER, H. C., BRUINING, H., and ATEN, A. H. W., 1947, *Philips Res. Rep.*, **2**, 171.

HEADRICK, L. B., and LEDERER, E. A., 1936, *Phys. Rev.*, **50**, 1094.

HERRMANN, G., and WAGENER, S., 1951, *The Oxide Coated Cathode* (London: Chapman and Hall), p. 271.

HILSCH, R., 1932, *Z. Phys.*, **77**, 427.

JACOBS, H., 1946, *J. Appl. Phys.*, **17**, 596.

JACOBS, H., and DOBISCHEK, D., 1951, *Phys. Rev.*, **81**, 1019.

JACOBS, H., MARTIN, J., and BRAND, F., 1952, *Phys. Rev.*, **85**, 441.

JACOBS, P. W. M., and TOMPKINS, F. C., 1952, *Proc. Roy. Soc.* A, **215**, 254, 265.

JONKER, J. L. H., 1947, *Philips Res. Rep.*, **2**, 331.

KÖNIG, H., 1951, *Z. Phys.*, **130**, 483.

MARTIENSSEN, W., 1952, *Z. Phys.*, **131**, 488.

MARTIENSSEN, W., and POHL, R. W., 1952, *Z. Phys.*, **133**, 153.

MATHESON, R. M., and NERGAARD, L. S., 1951, *R.C.A. Rev.*, **12**, 258.

METSON, G. H., 1949, *Proc. Phys. Soc.* B, **62**, 589.

MOORE, G. E., and ALLISON, H. W., 1950, *Phys. Rev.*, **77**, 246.

RUDBERG, E., 1936, *Phys. Rev.*, **50**, 138.

TAFT, E. A., and DICKEY, J. E., 1950, *Phys. Rev.*, **78**, 625.

TYLER, W. W., 1949, *Phys. Rev.*, **76**, 1887.

TYLER, W. W., and SPROULL, R. L., 1951, *Phys. Rev.*, **83**, 548.

DE VORE, H. B., and DEWDNEY, J. W., 1951, *Phys. Rev.*, **83**, 805.

WAGNER, E., 1931, *Trans. Electrochem. Soc.*, **59**, 223.

WOODS, J., and WRIGHT, D. A., 1952, *Brit. J. Appl. Phys.*, **3**, 323; 1953, *Ibid.*, **4**, 56.

An Experiment on the Radial Motion of Charged Particles from a Pulsed Arc Discharge

By N. L. ALLEN*
Electron Physics Department, University of Birmingham

Communicated by J. Sayers; *MS. received* 2nd *April* 1953,
and in amended form 6th *August* 1953

Abstract. A systematic investigation has been made of the currents drawn to a metal cylinder surrounding a concentric high-current arc in hydrogen in the presence of a longitudinal magnetic field. Characteristic curves show currents drawn with various positive and negative potentials applied to the cylinder. The shape of the curves is discussed qualitatively in terms of the radial motion of charged particles in the electric and magnetic fields. Consideration is given to the possibility that the action of the arc current, with the magnetic field and electric field due to the cylinder, might produce an appreciable evacuation of gas molecules from the region around the arc column.

§ 1. INTRODUCTION

THE work to be described was undertaken as part of a programme of investigation of the effect of a longitudinal magnetic field on a heavy-current arc at low pressure. Earlier workers (e.g. Cummings and Tonks 1941) have shown that, in general, such a magnetic field has a constrictive effect on the arc column, due to the restricted motion of charged particles outwards from the arc across the lines of force. Some confusion has arisen on this point, however, due to the various conditions and tube radii employed. These difficulties have recently been resolved (Champion 1952) and it is now considered that a magnetic field constricts an arc, provided the diameter of the containing vessel is greater than the natural diameter of the arc column at the pressure employed.

It might be expected that where the gas pressure is low, and the arc current very high, an appreciable proportion of the neutral molecules which pass into the column in their thermal motion would become ionized and, in the absence of extensive recombination, be held there by a strong magnetic field. As a result, a vacuum would develop around the positive column which would itself be expected to have a reduced potential gradient. It was felt that the use of a conducting cylinder surrounding the column might assist in producing this 'pumping effect'. Thus, if the cylinder were given a positive potential relative to the plasma, outwards diffusion of positive ions would be checked by both the electric and magnetic fields, while the electrons, though attracted outwards by the electric field, might be held in by the magnetic field. A reduction in pressure of the neutral molecules outside the column would further impede the outward flow of ions and electrons, if it is assumed that collisions are chiefly responsible

---

* Now at the Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Mass., U.S.A.

for their motion across the field. In the extreme case, the process leads to a picture of complete cut-off of current to the cylinder when all the gas molecules are held in the arc column.

Direct pressure measurements were very difficult under conditions of pulsed arcs in a confined space, since the arc itself interfered with the normal operation of conventional measuring devices. Therefore, a systematic study was made of the currents flowing to the cylinder itself across the magnetic field, in the hope of obtaining an indication of any pressure change which might occur. In addition, comparison was made with the work of Bohm, Burhop, Massey and Williams (1949), who employed arcs of lower current in an analogous arrangement.

§ 2. Experimental Technique

The discharge tube is shown in fig. 1. A cold-cathode arc was initiated between tungsten rod electrodes on the axis of an open-ended nichrome cylinder of diameter 3·2 cm and length 10 cm. The axial electrodes were 5 cm apart and symmetrically placed. Electrical contact with the cylinder was made through the supports S. The tube was evacuated and outgassed by induction heating of the cylinder and prolonged passage of discharges. Hydrogen was used in the experiments at pressures of $1{\cdot}5 \times 10^{-2}$ mm, 0·4 mm and 2·0 mm Hg.

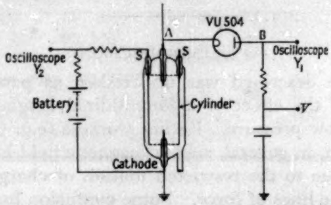


Fig. 1. The connections to the discharge tube. The shielding and solenoid around the tube are omitted.

The method used for initiation of the discharge at the lowest pressures has been described in a previous paper (Allen 1952), and it is necessary to state only that a rectangular pulse of current of amplitude 96 A lasting for 2 msec, passed through the gap at desired intervals, at a known instant in a pulsed magnetic field of 10 msec duration, and sinusoidal pulse shape (Champion 1950). The field was set up in a solenoid, with the discharge tube along its axis. The magnetic field was very nearly uniform along the length of the arc, falling off at the electrodes to about 90% of the value at the centre, and was used at various values up to 18 900 gauss.

The tube was screened electrostatically from the sharp change in potential which occurred at the solenoid terminals at the beginning of the magnetic field pulse. Tests showed that eddy currents in the nichrome cylinder were not serious, and caused little disturbance in the observed voltages on the cylinder, provided the resistance between cylinder and earth was not too high. The potential fall across the axial discharge was measured by noting the change in potential at B (fig. 1). Point A was the anode of the discharge tube which, prior to breakdown, was at high positive potential. A steady voltage V (about 300 v) was applied to the anode of the valve VU 504 which conducted during the discharge, when the potential at A fell below V. All measurements were made by means of an oscilloscope.

Various steady potentials were applied to the cylinder by means of batteries, in series with a resistance as shown in the diagram. The battery voltages did not exceed 300 v, positive or negative, and no discharge took place between the cylinder and the cathode in the absence of the longitudinal discharge pulse.

The static gas pressure in the discharge tube was measured with an ionization gauge, joined by wide tubes. Electrons from the gauge were restrained from passing into the high voltage gap and initiating unwanted discharges there by placing the poles of a magnetron magnet astride the connecting tube.

§ 3. EXPERIMENTAL RESULTS

Measurements were made at three pressures, $1{\cdot}5 \times 10^{-2}$, 0·4 and 2·0 mm Hg. The arcs appeared to fill the volume of the cylinder at zero magnetic field at pressures below 2 mm Hg. It may be seen that in this arrangement the arc was almost completely enclosed by the cylinder, so that few electrons and ions were lost to the glass walls of the tube. There was no evidence that arcs passed from the cathode to the cylinder and cylinder to anode, rather than in the longitudinal direction. Comparison with earlier work, using an arc enclosed only by the glass walls of a similar tube showed that the cylinder had no visible effect on the arc. The arc voltage was of the order of 80 v at the two higher pressures with zero magnetic field. It was not possible to pass arcs at $1{\cdot}5 \times 10^{-2}$ mm Hg without the magnetic field.

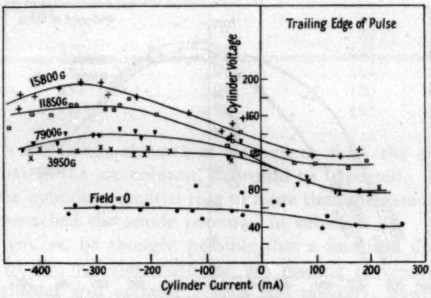


Fig. 2. Cylinder voltage–current curves for pressure of 2 mm Hg.

Preliminary readings taken with the cylinder (*a*) 'floating', (*b*) connected to cathode with a resistance and (*c*) at positive and negative potentials relative to cathode showed that different electrical connections to the cylinder had no appreciable effect on the voltage–current characteristics of the arc.

Systematic observations were carried out at the three pressures on the electrical characteristics of the cylinder during arc pulses and at various values of the axial magnetic field, when various steady positive and negative potentials were applied to the cylinder. Three typical sets of experimental curves summarizing the trend of the results are given in figs. 2–4. The convention employed is that electron current from the battery to the cylinder is regarded as negative.

The curves show that at 2 mm pressure, the pulse cylinder voltage appeared asymptotic to certain values when the current in either direction was large. The same was true at 0·4 mm pressure, but at $1{\cdot}5 \times 10^{-2}$ mm Hg a different form was shown (fig. 4) with no tendency to be asymptotic to a definite cylinder voltage. In fact, with large negative currents, the cylinder voltage was negative. This result was reasonable, if the impedance between the cylinder and earth was small compared with that of the external circuit. Since increasing positive and negative currents corresponded to increasing applied battery voltages, the nature of the curves indicates that the impedance between the cylinder and the arc was reduced as the current increased.

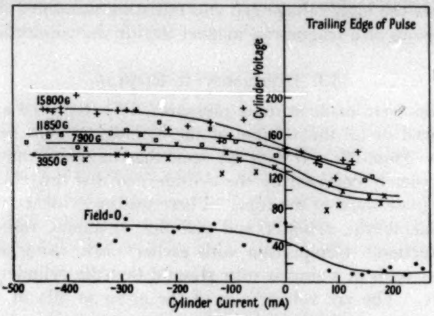


Fig. 3. Cylinder voltage–current curves for pressure of 0·4 mm Hg.

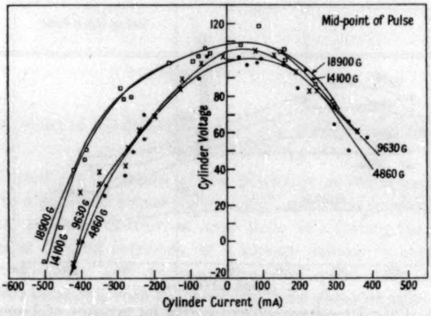


Fig. 4. Cylinder voltage–current curves for pressure of $1{\cdot}5 \times 10^{-2}$ mm Hg (some experimental points are omitted for clarity).

A notable feature of the curves at 2 mm and 0·4 mm pressure was that at all settings of the magnetic field, the decrease in cylinder voltage in passing from negative to positive current was nearly constant, though increasing a little at higher magnetic fields.

At constant pulse cylinder current, the voltage showed a large increase with magnetic field at 0·4 and 2·0 mm. The largest increase occurred between fields of 0 and 4860 gauss, the rise after that being less marked for further equal increases in magnetic field. The results at $1{\cdot}5 \times 10^{-2}$ mm Hg show that the effect of increases in magnetic field was smaller than at other pressures, but the increase over the zero-field case was unknown.

A short series of measurements was made at 2·0 and 0·4 mm pressure with the cylinder connected to earth (cathode) through a monitoring resistance of 110 kΩ, but unconnected to any other source of potential. The results are given in table 1. In this arrangement, the current to the cylinder was very small (less than 1 mA) and the pulse cylinder voltage was determined by the longitudinal arc itself. Thus there were, in effect, two sets of measurements of pulse voltage for the cylinder for zero current—that determined from the curves with the applied voltage adjusted to give zero net current, and that with a large resistance between the cylinder and earth.

Table 1. Results showing the pulse cylinder voltages when the cylinder is connected to earth through a 110 kΩ resistance.

(1) *Pressure* = 2 *mm Hg*

Field (gauss)	Arc voltage	Arc current (A)	Cylinder voltage	$\frac{\text{Cyl. voltage}}{\text{Arc voltage}}$
0	80	106	36	0·45
4860	175	97	86	0·49
9630	190	97	98	0·52
14100	226	96	122	0·54
18900	234	84	128	0·55

(2) *Pressure* = 0·4 *mm Hg*

Field (gauss)	Arc voltage	Arc current (A)	Cylinder voltage	$\frac{\text{Cyl. voltage}}{\text{Arc voltage}}$
0	75	105	39	0·52
4860	178	97	143	0·80
9630	181	96	169	0·93
14100	192	94	170	0·89
18900	206	90	195	0·95

In the latter case, in the absence of a magnetic field, the cylinder took up the mean potential of the arc column, as would be expected. With increasing magnetic field, the cylinder potential rose to more than the mean potential of the column, and approached the anode potential in value.

It might, of course, be thought possible that a localized discharge was set up at the ends of the cylinder, with the arc passing between the anode and cylinder, and cylinder and cathode. This was unlikely, however, for in the magnetic field the pulse cylinder voltage approached the arc anode voltage to a point where it would be impossible for the mechanisms of a discharge to operate.

It may be added that the passage of each arc resulted in a decrease of pressure, due to 'clean-up' of about 20% at $1{\cdot}5 \times 10^{-2}$ mm Hg. This was recorded by the gauge, but due to the relatively large distance between the discharge and the gauge (about 20 cm), the latter was insensitive to any detailed pressure fluctuation during the arc pulse.

§ 4. DISCUSSION

4.1. *The Shape of the Curves at Zero Magnetic Field*

As already indicated, the floating potential of the cylinder for all pressures lies between the cathode and anode potentials. On either side of this point there extends a current–voltage characteristic which for small voltage changes is linear, and is equivalent to a differential resistance of 200 ohms at a pressure of 0·4 mm. This resistance varies with pressure from 125 ohms at 2 mm to 250 ohms at $1{\cdot}2 \times 10^{-2}$ mm. For larger positive or negative voltages, at the higher end of the pressure range, the differential resistance between cylinder and arc is zero, and indicates the onset of a local discharge between the cylinder and arc.

4.2. *The Shape of the Characteristics in the Magnetic Field*

The magnetic field raised the voltage required to maintain a given current to the cylinder. At 2 mm and 0·4 mm pressure, the largest increase occurred with the relatively small increase of field from 0 to 4860 gauss. The effect of further increasing the field was smaller. Comparison may be drawn here with the case of a *heavy* current discharge between coaxial cylinders (Allen 1951) in which a similar 'saturation' was observed. At $1{\cdot}5 \times 10^{-2}$ mm the value of the magnetic field made relatively little difference to the pulse cylinder voltage. Visually, the constriction in the arc brought about by the field was most marked for the initial increase of magnetic field, though at a pressure of $1{\cdot}5 \times 10^{-2}$ mm the constrictive effect of fields of all strengths was slight. This was a result of the large 'natural radius' of the arc at this pressure (cf. § 1). The increase in arc anode voltage brought about by the field was greatest for the initial increase in field strength, rising less sharply with further increase.

An increase in the cylinder voltage required to maintain the same current with decreasing pressure would be expected with an increase of magnetic field, since electronic and ionic diffusion across the field depend upon collisions with gas molecules. However, there is evidence that electron diffusion across a magnetic field is aided by the plasma fluctuations set up by the magnetic field in the discharge. In these experiments the noise level on both the cylinder voltage and the arc current and voltage wave forms was as great at $1{\cdot}5 \times 10^{-2}$ mm as at the higher pressures. In general, the noise increased from a very small amount with zero magnetic field to a high level with the first increase of field, but showed little further increase at higher fields. This might offer some explanation of the very slight differences between the curves at $1{\cdot}5 \times 10^{-2}$ cm, but the fact that the field produces little constriction of the main arc is also very important.

4.3. *Comparison with the Work of Bohm et al.*

This work may be compared with that of Bohm, Burhop, Massey and Williams (1949) who studied the currents to the walls of a rectangular carbon box surrounding an arc in a gas at low pressure. The arc plasma was produced in the carbon box by electrons accelerated from a hot filament by a simple form of electron gun.

The results showed that the wall current was reduced by an increasing magnetic field, though the reduction compared with the zero-field case was not indicated. At the higher magnetic fields a degree of 'saturation' is discernible analogous to that observed by the present writer. Their results confirm that the magnetic field, while having a large effect on the current flow in increasing from zero to a comparatively small value, has a smaller effect when rising to larger values. The phenomenon is general for the cases of magnetic fields applied longitudinally and transversely to discharges. These conclusions are drawn from experiments with the longitudinal and transverse arcs with heavy currents, as well as in the work of Bohm *et al.* with lower currents transverse to the field.

Bohm *et al.* repeated their experiment at different pressures, keeping all other quantities constant. It was found that the wall current increased with increasing pressure under conditions where the relative degree of ionization was small, but where this was larger, the increase of wall current with pressure was less rapid.

These facts may be correlated with the present experiments in the following way: The curves of pulse cylinder voltage against current show that a large increase in voltage is required to maintain a given current for the first increase in magnetic field, but smaller increases are necessary with further equal increases in magnetic field. A comparison of the changes in the transverse direction with those in the longitudinal direction shows the relative effect on the magnetic field in the two directions. Use is made of some results for the positive column gradient by Dr. K. S. W. Champion, at various magnetic fields for arcs of current 96 A passing in the longitudinal direction. These results show that the positive column gradient tends to decrease in an increasing magnetic field.

To obtain a measure of the impedance existing between the cylinder and arc plasma, the value of half the total arc voltage is subtracted from the cylinder voltage. Then the expression

$$\frac{\text{Cylinder voltage} - \text{half arc voltage}}{\text{Positive column gradient}}$$

gives an indication of the ratio of the impedance to electron motion in the transverse and longitudinal directions for the conditions of the experiment. The values of this ratio are given in table 2. A constant cylinder current is not being dealt with in this case, but this is not a serious complication, since the shape of the curves is similar.

Table 2. Values of the ratio $\dfrac{\text{Cylinder voltage} - \text{half arc voltage}}{\text{Positive column gradient}}$ at two instants during the pulse: (*a*) at the mid-point in time and (*b*) at the end.

(*a*) *Mid-Point of Pulse*

Field (gauss)		0	4860	9630	14100	18900
Ratio	(i) 2 mm pressure	1·05	2·12	1·87	5·4	8·0
	(ii) 0·4 mm pressure	1·02	9·4	12·8	19·2	22·0

(*b*) *Trailing Edge of Pulse*

Field (gauss)		0	3950	7900	11850	15800
Ratio	(i) 2 mm pressure	0·63	2·9	4·65	8·7	10·1
	(ii) 0·4 mm pressure	0·29	6·5	8·8	13·8	14·2

It is seen that the value of the ratio increases with an increasing magnetic field. The impedance parameter across the field may be taken as that characteristic of a plasma in a transverse magnetic field, without the complication of a large cathode fall of potential. At the cylinder surface, a sheath will be formed, the nature of which will vary along its length. The current to the cylinder also varies from point to point, but in the presence of a field, the positive ion current density will be greater, in proportion to the electron current density, than is the case at zero magnetic field.

The results also show clearly the increase in impedance in the transverse direction brought about by the reduction in pressure from 2 to 0·4 mm Hg. Results for the positive column gradient at $1{\cdot}5 \times 10^{-2}$ mm Hg are not available, but from the nature of the curves the same trend would be expected, though with a greater saturation effect than at the higher pressures. The ratios in table 2 indicate that the field offers a smaller impedance to the flow of electrons across the lines of force than may be calculated assuming progress depending on simple collisions alone.

Diffusion to the cylinder may be aided by the existence of plasma oscillations in the positive column (Backus 1949). Bohm (1949) suggests that if the equipotential surfaces imagined to surround the arc are symmetrical about the arc column (as is reasonably assumed in a symmetrical configuration such as the present one), then there is no stable mode of getting electrons and ions away to the walls, except by collision. This process alone would be exceedingly slow in the magnetic field, and a piling up of ionization would occur, leading to instability. Such a condition would be expected to result in plasma oscillations in whose fields the ions and electrons would drain to the walls, setting up a dynamic equilibrium again.

§ 5. CONCLUDING REMARKS

The results of this work confirm earlier work in showing that the impedance offered by the action of the transverse magnetic field to a discharge current in a gas shows a saturation effect at the higher magnetic fields. Within the limits of accuracy of the experiment, the effect appeared to become stronger as the pressure was reduced. These facts are taken as an indication that no appreciable cut-off of current from the arc to the cylinder occurred as the result of a strong 'pumping action' by the arc in the magnetic field. The change in saturation effect with pressure is also contrary to the expectation, from simple kinetic theory, that a pumping action would be most likely at the lowest pressures, but this may be a result of the condition of the experiment, since the natural radius of the arc is increasing at the lower pressures.

A possible reason for the apparent absence of a pumping effect is that neutral molecules which are ionized in the arc column may recombine extensively at the walls near the cathode end of the tube, setting up an appreciable pressure gradient which returns gas molecules to the body of the tube. Thus a circulation of gas may occur in the chamber, masking the effects looked for by this method. The oscilloscopic method of observation shows that the electrical behaviour of the cylinder and arc is substantially the same throughout the pulse and characteristic of approximately uniform conditions in time. It has been noted also that the arc voltage drop remains unaltered, within the limits of accuracy, for all battery voltages applied to the cylinder. This may indicate that the cylinder has little effect upon the arc, though it must be remembered that the increased molecular agitation in the arc column may itself be comparable with any postulated increase in pressure due to a pumping effect. Thus such a measurement is regarded as insensitive. Further work, employing more refined techniques, such as the use of a radioactive 'tracer' in the gas, would be necessary to yield more information.

ACKNOWLEDGMENTS

The writer is grateful to Professor J. Sayers, who suggested this experiment, for his advice at all stages of the work, and also to Dr. K. S. W. Champion for discussion. A grant from the Department of Scientific and Industrial Research is also acknowledged.

REFERENCES

ALLEN, N. L., 1951, *Proc. Phys. Soc.* B, **64,** 276; 1952, *Ibid.*, **65,** 697.

BACKUS, J., 1949, *Characteristics of Electrical Discharges in Magnetic Fields*, ed. A. Guthrie and R. K. Wakerling (New York : McGraw–Hill), chap. 2, p. 353 ff.

BOHM, D., 1949, *Characteristics of Electrical Discharges in Magnetic Fields*, ed. A. Guthrie and R. K. Wakerling (New York : McGraw–Hill), chap. 4, p. 87 ff.

BOHM, D., BURHOP, E. H. S., MASSEY, H. S. W., and WILLIAMS, R. M., 1949, *Characteristics of Electrical Discharges in Magnetic Fields*, ed. A. Guthrie and R. K. Wakerling (New York : McGraw–Hill), chap. 5, p. 107 ff.

CHAMPION, K. S. W., 1950, *Proc. Phys. Soc.* B, **63,** 795; 1952, *Ibid.*, **65,** 329.

CUMMINGS, C. S., and TONKS, L., 1941, *Phys. Rev.*, **59,** 514.

A New Microscope Principle

By J. M. COWLEY
Chemical Physics Section, Division of Industrial Chemistry, Commonwealth Scientific and Industrial Research Organization, Melbourne, Australia

Communicated by I. W. Wark; MS. received 3rd June 1953

Abstract. A high-resolution image may be derived from a large number of 'dark-field' images of normal resolution, obtained by varying the angle of incidence of the electron beam in a standard electron microscope. The intensity distribution in each dark-field image is multiplied by an appropriate cosine function. Summing gives the function, for one dimension,

$$C(x) = \int V(X)\, G(-X)\, V(X+x)\, G(x-X)\, dX$$

where $V(x)$ is the potential distribution in the specimen and $G(x)$ defines the loss of resolution due to lens imperfections in normal micrographs. From this, $V(x)$ may be derived. The practical applications of the method are limited by requirements of specimen and instrumental stability.

THE smallest distance which can be resolved by present day electron microscopes is of the order of 10 Å. This is very much greater than the limit of resolvable distance imposed by the wavelength of the electrons used (about 0·05 Å). Apart from problems of specimen stability, the main factor preventing any further improvement of the resolving power is the limitation of the effective aperture of the objective lens by spherical aberration.

There is a strong incentive for attempts to improve the resolving power of the electron microscope, in that a ten-fold improvement, say, would allow individual atoms to be resolved, and the main features, if not the details, of the atomic arrangement within molecules or crystals to be determined. Since it does not seem possible that the resolving power may be increased markedly by the redesigning or correction of the objective lens, elaborations of the microscope technique are needed to take advantage of all possible properties of the electron beam and allow a high resolution image to be derived from images obtained with imperfect lenses. Gabor (1949) has proposed an ingenious scheme in which a sufficiently coherent electron source is used, and the electrons elastically scattered within a certain solid angle are recombined with the transmitted electron beam to form a 'hologram' from which an image is reconstructed by a light-optical system. Limited success has been achieved in the employment of this method (Haine and Mulvey 1952).

Another method of increasing the effective resolving power is to make use of the many 'dark-field' images of the specimen obtained from beams of elastically scattered electrons. With the incident electron beam along the electron-optical axis of the microscope, the elastically scattered, or diffracted, electrons diverge from the specimen and either are stopped by apertures or lens pole-pieces or else, because of the spherical aberration of the objective lens, form out-of-focus 'dark-field' images, displaced from the image formed by the transmitted electrons.

If, however, the incident beam is tilted at an angle to the microscope axis, one of these scattered beams may be made to travel down the electron-optical axis and form an undisplaced, well-focused dark-field image. By varying the tilt of the incident beam, a large number of such dark-field images may be recorded. These may be combined in such a way that a single image of greatly improved resolution may be derived.

This principle could be applied to microscopy with light, x-rays or ions, as well as with electrons, but since the application to electron microscopy seems most practicable and would, if successful, extend the total range of microscopic methods, only this case will be considered.

We consider for simplicity a one-dimensional object with a potential distribution $V(y)$. We make the assumption that the scattering is purely elastic and absorption is negligible, so that the amplitude of the wave forming the image may be represented by a real function. A more rigorous treatment using a complex transmission function is possible, but gives essentially the same result. The diffraction pattern in the back-focal plane of the object lens is $E(\xi)$, the Fourier transform of $V(y)$. The image is given by the inverse transform as $V(x)$, where $x=-y$ except for a scale factor. In an electron microscope, if the diffraction pattern in the back-focal plane is limited by an aperture, assumed symmetrical, the image is the Fourier transform of $E(\xi)f(\xi)$, where $f(\xi)$ is the 'aperture function':

$$\mathscr{F}^{-1}E(\xi)f(\xi)=\overgroup{V(x)\,F(x)}\equiv\int V(X)\,F(x-X)\,dX.$$

Here $F(x)$ is the Fourier transform of $f(\xi)$, and the 'fold' or 'faltung' of V with F represents the loss of resolution.

If now the beam incident on the object is tilted so that it is parallel to the initial direction of the diffracted beam corresponding to $\xi=\eta$, the effective portion of the diffraction pattern is given by $E(\eta-\xi)f(\xi)$, i.e. tilting of the incident beam translates the diffraction pattern in the back-focal plane so that, instead of the part of the diffraction pattern around $\xi=0$, the part around $\xi=\eta$ passes through the aperture. The 'dark-field' image produced is then

$$V'(x,\eta)=\mathscr{F}^{-1}E(\eta-\xi)f(\xi)=\overgroup{[V(x)\exp(-2\pi i\eta x)]\,F(x)}.$$

The aberrations of the lenses will result in a further loss of resolution, so that the image obtained is

$$V(x,\eta)=\overgroup{V'(x,\eta)\,H(x)}=\overgroup{[V(x)\exp(-2\pi i\eta x)]\,G(x)},$$

where $G(x)=\overgroup{F(x)\,H(x)}$. The recorded intensity of the image is proportional to

$$|V(x,\eta)|^2=V(x,\eta)\,V^*(x,\eta).$$

If one particular point of the image, say $x=0$, is chosen, the intensity at this point is given as a function of η by

$$\begin{aligned}I(0,\eta)&=[V(x,\eta)\,V^*(x,\eta)]_{x=0}\\&=\iint V(X)\,\mathrm{e}^{-2\pi i\eta X}\,G(-X)\,V(Y)\,\mathrm{e}^{2\pi i\eta Y}\,G(-Y)\,dX\,dY.\end{aligned}$$

The Fourier transform of this intensity function is

$$\begin{aligned}P'(x)&=\int I(0,\eta)\,\mathrm{e}^{-2\pi i\eta x}\,d\eta\\&=\int V(X)\,G(X)\,V(X+x)\,G(X+x)\,dX,\end{aligned}$$

which is the Patterson function of the limited region of the image given by $V(x)\,G(x)$.

The problem of obtaining $V(x)G(x)$ from $P'(x)$ is the well-known problem occurring in the structure analysis of crystals by x-ray and electron diffraction methods. For a two-dimensional image, if the half-width of $G(x)$, considered to approximate to an error function, is 10–20 Å, the problem would be of the same order as the problems commonly occurring in crystal structure analysis, and the methods of solution developed for use in structure analysis, such as the 'image-seeking' methods of Buerger (1951) and others, could be applied. Usually it is not possible to resolve such a Patterson function uniquely, particularly since the portion of image chosen will, in general, be non-centrosymmetric. The image could be deduced only in the most favourable cases.

If only the intensity at one particular point of the image is taken, however, much of the information available concerning the portion $V(x)\,G(x)$ of the image has not been used. One knows the intensities of the images at all points within the region defined by $G(x)$. In order to obtain a function containing more information than the Patterson function the intensity distribution of each image may be multiplied by a cosine function and the resulting functions summed. Thus the constant coefficients of the Patterson summation are replaced by functions of x.

For each η one multiplies the image intensity distribution by $\cos 2\pi\eta x$. Summing over all η, one obtains

$$C(x)=\int \overgroup{[V(x)\exp(-2\pi i\eta x)]\,G(x)}\,\overgroup{[V(x)\exp(2\pi i\eta x)]\,G(x)}\cos(2\pi\eta x)\,d\eta$$

$$=\iint V(X)\,G(x-X)\,V(Y)\,G(x-Y)\int \exp\{-2\pi i\eta(X-Y)\}\cos(2\pi\eta x)\,d\eta\,dX\,dY.$$

The integral over η reduces to the sum of two δ-functions, giving

$$C(x)=\int V(X)\,G(-X)\,V(X+x)\,G(x-X)\,dX.$$

The procedure for obtaining $C(x)$ in practice would probably be as follows. Dark-field images are recorded for a large number of tilts of the beam, so chosen that each portion of the diffraction pattern in the back focal plane, which has a ξ-value less than a certain maximum, passes through the aperture and so contributes to the image at least once. The intensity distribution for each dark-field image is then multiplied by the appropriate cosine function and the resulting functions are added together. These multiplications and the addition could be carried out by one of the several optical or electronic methods which appear feasible. The function $C(x)$, unlike $P'(x)$, is not, in general, centrosymmetric. As well as information on the vector separations of scattering points, it contains information on the positions of these scattering points with respect to the point chosen as origin. It should, then, be possible to obtain $V(x)\,G(x)$ from $C(x)$ more readily than from $P'(x)$.

Once $V(x)$ has been deduced over the limited field defined by $G(x)$, the knowledge of $V(x)$ in this field could be used to simplify the deduction of $V(x)$ in other, overlapping fields. In this way the 'field of view' could be extended, and also a portion of the specimen of known structure, such as a well-ordered crystal, could be used in the imaging of a neighbouring region of unknown structure.

Provided that the object is only a few atoms thick, an image with a resolution of 1 Å or better will consist of well-defined atomic potential maxima, which may be approximated by 'point functions'. In this case in particular $V(x)$ may be derived by considering the function

$$B(x)=P'(x)-C(x)=\int V(X)\,V(X+x)\,G(X)\,[G(X+x)-G(X-x)]\,dX,$$

where $G(x)$ is considered to be centrosymmetric and approximately an error function. If $V(x)$ has peaks at x_1 and x_2, the function $B(x)$ has peaks at (x_2-x_1) and (x_1-x_2).

Then for $x_2>x_1>0$, $B(x_2-x_1)$ is positive and $B(x_1-x_2)$ is negative;

for $x_2>0>x_1$, $B(x_2-x_1)$ and $B(x_1-x_2)$ are negative;

for $0>x_2>x_1$, $B(x_2-x_1)$ is negative and $B(x_1-x_2)$ is positive.

If $x_1=0$, $B(x_2-x_1)=0$; if $x_2=0$, $B(x_1-x_2)=0$.

Hence, if the origin of $V(x)$ is chosen at the centre of one atom, all the peaks in $B(x)$ corresponding to vectors from this atom to other atoms vanish, provided that the interatomic vectors occur only once in the range defined by $G(x)$. Thus the positions of all atoms may be determined by observing the behaviour of the vector peaks in $B(x)$ as the origin of coordinates is moved with respect to the image.

Alternatively, if the function $G(x)$ is known, the coordinates x_1 and x_2 of two maxima may be derived from the known value of x_2-x_1 and the relation

$$\frac{B(x_2-x_1)}{B(x_1-x_2)}=\frac{G(x_1)\,[G(x_2)-G(2x_1-x_2)]}{G(x_2)\,[G(x_1)-G(2x_2-x_1)]}.$$

For the practical realization of this method of microscopy a conventional-type electron microscope need be modified only by the incorporation of a device to allow the variation of the angle of incidence of the electron beam on the specimen, such as a combination of electrostatic deflecting plates and a magnetic lens. No special requirements in the way of beam coherence are involved. It is sufficient that the divergence of the beam should be such that the resolution in the diffraction pattern, $E(\xi)$, is appreciably better than the half-width of $f(\xi)$. This is readily achieved.

In principle, the resolution obtainable by this method is limited only by the wavelength of the electrons. However, the number of dark-field images which must be recorded in order to resolve c Å in a two-dimensional image is of the order of n^2, where n is the ratio of the half-width of $G(x)$ to c, i.e.

(resolution of each image)/(ultimate resolution).

Each image would have to be recorded under the instrumental conditions of magnification, instrument and specimen stability which would be required to obtain an image with c Å resolution if perfect lenses were available. The point taken as the zero point in the calculation of $C(x)$ and $B(x)$ must be identified in each image with an accuracy of better than c Å. This can only be done if the zero point of each image comes to the same point on the viewing screen or recording plate. Hence, over the period required to record all the dark-field images, the specimen must not move more than c Å, and the magnification of the lenses, stray fields, and so on, must not vary enough to move the image by more than the equivalent of c Å. With the high magnifications required, the time necessary for the recording of each dark-field image may be of the order of a minute. The total period for which the conditions of extreme stability of the specimen and electron-optical system must be maintained would probably be of the order of several hours if the resolution necessary for the imaging of atoms was to be achieved. Such stability requirements are well beyond those of present-day electron microscopes, and present serious experimental difficulties which must be overcome before this method can be used to give any very marked improvement in resolution.

The derivation of the image from the function $C(x)$ or $B(x)$ is straightforward only when the image may be approximated by a set of peak-functions, i.e. when individual atoms or groups of atoms are well resolved. In other cases, where the image can be described only in terms of a continuous distribution of potential, the derivation of the image may be possible if $C(x)$ or $B(x)$ and the image are approximated by weighted sets of points on a fine grid in the same way as are the Patterson and Fourier projections calculated by summing Fourier series in normal crystal-structure analysis.

REFERENCES

BUERGER, M. J., 1951, *Acta Cryst., Camb.*, **4**, 531.
GABOR, D., 1949, *Proc. Roy. Soc.* A, **197**, 454.
HAINE, M. E., and MULVEY, T., 1952, *J. Opt. Soc. Amer.*, **42**, 763.

Precipitation Processes in Copper–Iron Alloys

By J. REEKIE, T. S. HUTCHISON and F. E. HETHERINGTON
Royal Military College of Canada, Kingston, Ontario, Canada

MS. received 29th June 1953, and in amended form 14th September 1953

Abstract. Measurements of the electrical resistivity of copper–iron alloys, in the annealed condition, have shown that the concept of 'coherent' and 'incoherent' precipitate can explain the observed features in the resistivity curve.

Cold working of the alloys has been shown to result in resistivity changes which are compatible with the idea of 'locking-in' of dislocation lines about the precipitated particles of iron. Calculation from x-ray observations of the distance apart of these particles gives a figure of approximately 10^{-5} cm. This is quite close to that estimated from Orowan's equation for the distance between precipitate particles most effective for locking-in of dislocations in an age-hardened alloy.

Estimates of the number of vacancies present in a 95% cold worked alloy suggest that vacancies alone will not account for the observed resistivity increases. At high degrees of cold working, dislocations are probably the predominant factor in increasing resistance.

§ 1. INTRODUCTION

The physical properties of metallic solid solutions generally undergo relatively large changes as soon as precipitation of the solute atoms commences. Such changes are perhaps best known among the many 'age-hardening' alloys. In these, depending on the temperature at which ageing is carried out, more or less rapid variations are observed in such properties as mechanical hardness, ductility and strength, lattice parameter, electrical and thermal resistance, magnetic coercive force and remanence. In many cases simultaneous changes occur in two or more of the above properties; for instance, parallel changes in mechanical hardness and electrical resistivity are frequently found. Prior mechanical treatment also may have a considerable influence on the ageing processes, even in one alloy. Furthermore, the ageing curve is rarely in the form of a simple increase or decrease in the observed property. The copper–beryllium alloys for example, over a wide range of ageing temperatures, show an initial increase in hardness followed by a slower decrease as ageing proceeds.

In order that ageing should be observed in an alloy it is a prerequisite that the solubility of one component should decrease with decreasing temperature. Only then, of course, can precipitation from the supersaturated solid solution take place. Since this fact was realized many workers have investigated the factors influencing precipitation, particularly in its initial stages, and have attempted to correlate ageing phenomena with precipitation processes. Geisler (1951) has discussed this problem at considerable length, and has extended previous ideas in order to explain many of the ageing curves observed in a large number of alloys. Essentially these explanations assume the presence of 'coherent' precipitate in the early stages, followed by the formation of 'incoherent' precipitate as ageing progresses.

In general the presence of coherent precipitate results in 'coherency strains' in the alloy, which in turn modify its physical properties. These strains are relieved, as ageing proceeds, by the breaking away of the precipitate into incoherent form, with resulting further changes in physical properties. It will be obvious that, since depletion of solute atoms from the matrix also markedly affects the physical characteristics of an alloy, the overall ageing curve in any particular case can be quite complex.

Ageing can, of course, be followed by observing changes of any physical property as a function of time, and measurements of mechanical hardness have been most frequently used for this purpose. However, the electrical resistivity is, under certain conditions, a more sensitive indicator of the internal state of an alloy and many isothermal ageing processes have been followed using resistivity measurements.

In order to see if the 'coherency–incoherency hypothesis' can equally well explain effects due to compositional changes, we have in the present work investigated the electrical resistivity of a number of alloys of copper with small amounts of iron, up to 1·4% by weight. The concept of a coherent precipitate existing over a certain range of iron content, this precipitate becoming predominantly incoherent at higher iron contents, does appear to be able satisfactorily to account for the variation of resistivity of the annealed alloys. The influence of cold working on resistivity has also been investigated, and these observations again lead to the conclusion that most of the iron must transform into incoherent form above a certain percentage. Direct observation by x-ray diffraction methods confirm the latter conclusion.

§ 2. Experimental Procedures

2.1. *Preparation of Alloy Samples*

For the resistivity measurements, nine alloys of copper and iron were obtained from the Metals Division of Imperial Chemical Industries Ltd., the iron content varying over the range from 0·04% to 1·34%. These alloys were originally in the form of hard-drawn rods approximately 5 mm in diameter; the respective iron contents and resistivities are given in the table. The 'pure' copper was supplied by Messrs. Johnson, Matthey and Mallory Ltd., the total metallic impurities quoted amounting to less than 0·001%. In order to have uniform size in the specimens they were all initially drawn down through circular hardened steel dies to a diameter of 3·5 mm, and cut to approximately 15 cm in length.

Specimen no.	1	2	3	4	5	6	7	8	9	10
Iron content (wt.%)	†	0·04	0·15	0·23	0·26	0·35	0·40	0·50	0·79	1·34
Resistivity* (ohm cm × 10^6)	1·751	1·817	2·877	3·443	3·488	3·303	3·258	3·243	3·321	3·501

* in annealed condition at 27·1° c. † not detectable.

A uniform heat treatment was then applied by holding all specimens at 875°c for 10 hours in a high vacuum furnace, each being contained in an open Vicor glass tube to avoid contamination. Thereafter they were allowed to cool slowly to room temperature in the furnace over a period of about 20 hours. The true solubility limit of iron in copper at room temperature is known to be exceedingly small, and the above heat treatment does not produce an equilibrium alloy.

Previous work (Hutchison and Reekie 1951) has shown that such a heat treatment results in about 0·25% of the iron remaining in solution at room temperatures, thus forming a supersaturated solid solution.

2.2. *Resistivity Measurements*

A comparator potentiometer method was employed to determine the resistance of a rod specimen between well-defined knife-edges when carrying a current of approximately 100 mA. The circuit used was basically similar to that originally described by Kapitza and Milner (1937). Certain modifications were incorporated in order to improve the accuracy, and split photocell amplification was added into the galvanometer circuit. Figure 1 shows the complete circuit, which is for the most part self-explanatory. The resistances R_3 and R_4, whose ratio was approximately 100 : 1, effectively multiplied the slide wire reading on the potentiometer by this factor, and hence allowed voltages of the order of 10^{-5} to be determined with an accuracy of 1 part in 1000 when using a type K2 potentiometer. The slight non-linearity introduced by adding this resistance $R_3 + R_4$ across the potentiometer slide wire was determined by a separate calibration of the output voltage from the potentiometer and resistance together.

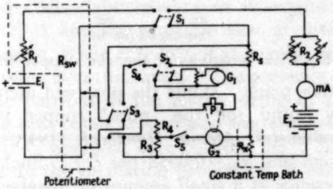


Fig. 1. Comparator potentiometer circuit used for measuring low resistances.

In order to obtain the specific resistance it is necessary to know the dimensions of the specimen, and it is easy to see that, particularly for small specimens, a large error in resistivity values can arise from inaccuracy in measurement of the specimen cross-sectional area. In the present work this was always obtained from a volume determination. Hydrostatic weighings on specimens of accurately known length gave density determinations directly and hence the cross-sectional areas. Observing all due precautions during weighings, this method gives cross sections with an accuracy an order better than any method involving direct measurement of the diameter.

Taking all known sources of possible inaccuracy into account the maximum error in the measured specific resistance is estimated to vary from 0·2% for the largest specimen (of 3·5 mm diameter) up to 1% for the smallest (of 0·76 mm diameter). Full details of the present method of measuring resistivity will be published elsewhere.

During potentiometer measurements specimens were maintained in a constant temperature bath at 27·1°c ± 0·02°c, and all specific resistivity figures refer to this temperature.

§ 3. EXPERIMENTAL RESULTS

3.1. *Resistivity of Annealed Alloys*

Figure 2 illustrates the variation of specific resistivity with percentage of iron in the copper–iron alloys after subjecting them to the heat treatment described previously. A relatively large increase in ρ is observed as the iron content increases from zero to 0·26%. Following this there is a smaller *decrease* up to 0·5% iron, and thereafter the resistivity again increases slowly and practically uniformly. As we have already shown, the heat treatment applied to these alloys results predominantly in the formation of a supersaturated solid solution, up to 0·26% iron content. Electrical resistivity would therefore be expected to increase in a linear fashion up to this point. While the observed increase is approximately linear, the resistivity found for the 'pure' copper is almost 2% higher than the accepted value (cf. *Metals Handbook*, 1948, p. 903, which gives $\rho_{20} = 1{\cdot}673 \times 10^{-6}$ ohm cm and $\alpha_{20} = 0{\cdot}0068$ per °c). Such a discrepancy could arise owing to the presence of a small amount of dissolved oxygen in the pure copper. On the addition of iron the oxygen would no longer affect the resistance appreciably, and the initial non-linearity of the resistance curve may well come about in this way.

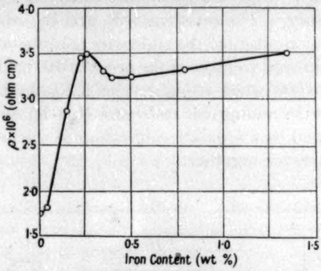


Fig. 2. Electrical resistivity of annealed copper–iron alloys.

In order to explain the shape of the resistivity curve for iron contents above 0·26% it is necessary to consider the manner in which precipitation processes can take place as the alloys cool slowly from 875°c. The solubility curve for iron in copper, according to Bauer and Hansen (1934), is shown in fig. 3. From this it appears that a temperature of 875°c would be sufficient to maintain nearly 1·5% of iron in solution. The highest iron content used here was 1·34%, so it may be assumed that initially all of the iron was in solution in the copper. Available data on diffusion coefficients show that a decrease in the rate of diffusion of the iron in

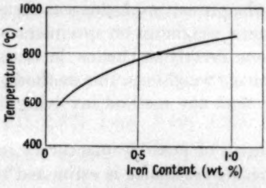


Fig. 3. Solubility of iron in copper at various temperatures (*after* Bauer and Hansen).

the copper matrix by approximately a factor of 10 is to be expected as the temperature falls from 875°c to 700°c (Smithells 1949). It may therefore be assumed that, as the alloys cool, the iron is able to diffuse relatively easily while the temperature still remains above about 750°c. From Bauer and Hansen's solubility curve this would imply that, once nucleation had been initiated, it would be relatively easy for the excess iron at any temperature to diffuse sufficiently to form discrete, stable particles from 1·34% down to about 0·45% iron. As will be discussed later, we have definite evidence that this is correct from the appearance of the body-centred cubic lines of iron in long exposure diffraction pictures of the alloys over the range 0·45 to 1·34%. These lines give the characteristic lattice spacing of iron, and the conclusion to be drawn from this is that over the above range the iron is necessarily in incoherent form, i.e. it is stable α-iron.

As the alloys cool still further the diffusion rate becomes rapidly lower, and at 650°c amounts to only about one hundredth the rate at 850°c. Consequently the iron still in solution becomes much less easily able to diffuse sufficiently to form free and unstrained precipitate particles. From 750°c, where the solubility limit amounts to 0·45% iron, down to 650°c, where the solubility limit falls to about 0·25%, it might be expected that the iron would therefore precipitate partially in a *coherent* form. Of course we have neglected any consideration of the activation energy necessary for the establishment of nuclei for precipitation. This factor must also be of importance in determining the nature of the precipitate. However, it appears probable that at 0·45% the precipitate is predominantly incoherent, while at 0·26% it is largely coherent, with very little iron remaining in solution over this range. The presence of coherent precipitate results in appreciable portions of the matrix being in a state of strain. The 'coherency strains' in turn result in increased scattering of conduction electrons, and the rise in resistivity observed here between 0·5% and 0·25% iron content is believed to be due to the increasing proportion of coherent precipitate over this range. Again we have x-ray evidence for the general correctness of this explanation; back-reflection Debye–Scherrer diffraction pictures show a diffuseness of the reflection spots from copper which is appreciably greater in the alloy containing 0·26% iron than in the 0·5% alloy.

By the time the alloys have cooled to about 650°c, at which the solubility limit is 0·25%, the diffusion is sufficiently slow to prevent aggregation of the iron. In those alloys containing less than 0·26% it therefore remains atomically dispersed while the cooling proceeds to room temperature and the resistivity follows the normal variation to be expected for a solid solution.

Where reliable data exist, it is generally found that the rate of diffusion of one metal in another becomes less as its concentration decreases. This would serve merely to accentuate the effect outlined above.

3.2. *Influence of Cold Working on Resistivity*

A great deal of information relevant to precipitation processes can be deduced from the behaviour of the electrical resistivity of copper–iron alloys when they are cold worked, and a short account of such resistivity measurements has already been given (Hetherington and Reekie 1951). Each of the nine alloys previously investigated, and the pure copper, were cold worked by drawing through circular steel dies. Drawing was always carried out in the same direction, at a constant rate of one foot per minute, and every precaution taken to avoid contamination of the specimens. The resistivity of each specimen was then measured as successively increasing reductions in area were made. Figures 4 and 5 show the results obtained; again all values refer to a temperature of 27·1°c. It will be clear that the curves fall roughly into two types; for all alloys above 0·25% iron the increase in resistivity on cold working follows much the same course and is very nearly of the same relative amount. Below 0·25% the resistivity change is not only much less, but is also markedly different in the different alloys. In the case of pure copper an overall increase of some 2% of the annealed value is found at high degrees of cold working, in agreement with earlier work of Rutter and Reekie (1950). In the 0·04% alloy the curve follows the same pattern, but with a somewhat increased value of resistivity. The resistivity of the 0·15% alloy, besides being generally higher, exhibits a sharp increase beyond about 90% reduction of area. The same feature is evident in the 0·23% alloy, but at a much earlier point in the cold working. For alloys of 0·26% iron content and beyond, the resistivity increases rapidly immediately cold working commences and the increase is very nearly the same for all the alloys up to 1·34% iron content.

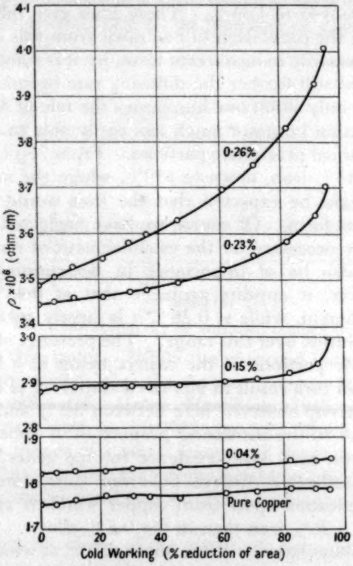


Fig. 4. Resistivity of cold worked copper–iron alloys (weight per cent iron content as shown).

This marked difference in behaviour, depending on whether the iron is held in solid solution or not, becomes of much more obvious significance if, instead of the resistivity ρ, we consider the difference between the resistivities of the alloy and of pure copper for equal degrees of cold working. This quantity, $\rho_{\mathrm{Fe,\,Cu}} - \rho_{\mathrm{Cu}}$, then measures the additional resistance resulting from the presence of the iron, above that of the copper matrix. It is evident that as soon as the iron is in precipitate form, the cold working becomes many times more effective in increasing the resistance of the alloy. This obtains in the range from 0·26% to 1·34% iron content, over which, as we have shown earlier, the iron is present either as coherent or incoherent precipitate, or as a mixture of both forms. Furthermore it is well known that cold working may initiate precipitation of the solute atoms in a supersaturated solid solution, so that below 0·26% iron one would expect a certain amount of initial cold working to be necessary *before* the rapid increase in resistivity began. This is evident in the 0·23 and 0·15% alloys, whereas it is to be presumed that in the 0·04% alloy even 95% cold working (reduction of area) is insufficient to initiate precipitation.

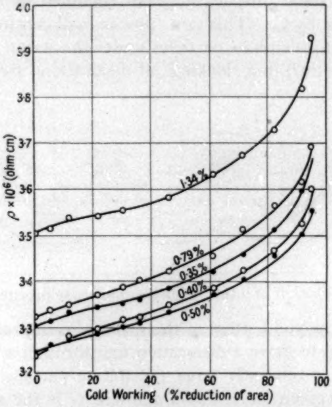


Fig. 5. Resistivity of cold worked copper–iron alloys (weight per cent iron content as shown).

It seems clear that when the iron is precipitated it is very much more effective in 'locking' the vacancies or dislocations (or, more generally, lattice disturbances) generated by cold working, and thus increasing the electrical resistance, than when it is atomically dispersed. In fact, before precipitation a large proportion of the generated vacancies and dislocations must become ineffective (presumably by mutual annihilation or absorption at grain boundaries). The solute iron atoms act merely as additional scattering centres, and this is reflected as a *constant* additional resistance above that of the pure matrix. Only when precipitation commences are additional dislocations maintained in the alloy. Of course, in order that the dislocations as they are generated should be 'locked-in' by the precipitate particles it is clearly necessary for these particles to be at a suitable distance apart. The quantitative aspects of this explanation will now be discussed.

3.3. *Average Distance between Precipitated Particles*

In order to account for the phenomena observed in precipitation hardening Nabarro (1948) and also Orowan (1948) have discussed the effect on yield stress of variation in size of the precipitate particles. Age hardening alloys in general show a maximum in the yield stress curve and it has been presumed that this maximum occurs at some critical particle size, the alloy being softer for both smaller and larger precipitate sizes. For a given total volume of precipitate the particle size determines the distance between particles, so we can equally well say that the yield stress maximum occurs when precipitate particles are a certain critical distance apart.

The physical process occurring when a dislocation line is driven by an externally applied stress towards a row of precipitate particles has been envisaged by Orowan (1948) to take place approximately as shown in fig. 6. Stress regions around the particles will tend to hold back the advancing dislocation line so that it will bulge forward between the particles, as at A. Increasing applied stress will cause the line to bulge still more, until it eventually breaks through the obstacles and at the same time leaves them encircled by small closed loops of dislocations, as shown by C. This same process will continue as new dislocation lines reach the row of precipitate particles under the action of the external stress, resulting, as will be clear, in a 'locking' of dislocations around the particles.

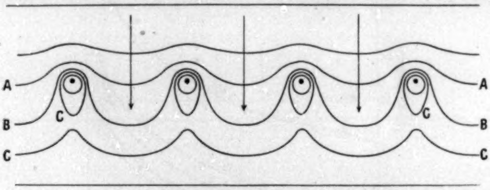


Fig. 6. The motion of a line of dislocations past precipitate particles.

On this picture, the yield stress of the material is approximately equal to the shear stress necessary to drive a dislocation line through a row of obstacles at a distance d apart. For relatively large precipitate particles Orowan (1948) has calculated this to be given by $Y=2Ga/d$ where G is the shear modulus and a the interatomic distance in the glide direction.

Nabarro (1948) has shown that for *small* values of d the yield stress is $Y=\text{const.}\sqrt{d}$. Clearly the curve of yield stress for an age-hardening alloy, if it is determined by these equations, must have a maximum at some value of d somewhat less than $d=2Ga/Y$. If we take values applicable to the present copper–iron alloys, viz. $G=6{\cdot}7\times10^{11}\,\text{dyn cm}^{-2}$, $a=2{\cdot}6\times10^{-8}\,\text{cm}$, and $Y=2{\cdot}2\times10^{9}\,\text{dyn cm}^{-2}$, we obtain for d a value of about $1{\cdot}6\times10^{-5}\,\text{cm}$. The optimum distance apart for 'locking' of dislocations to occur must be somewhat less than this and must therefore be about $10^{-5}\,\text{cm}$. It now remains to be determined whether the incoherent precipitate particles in our alloys can be about this distance apart.

We have taken long-exposure back reflection diffraction pictures of these alloys in the fully annealed state and have been able to obtain measurable iron lines in the case of all the alloys from 1·34% down to 0·5% iron. These lines correspond to a body-centred cubic structure having the normal iron spacing and therefore prove that the precipitated iron is in the normal stable α-phase. From measurements on the width of the iron lines estimates have then been made of the average size of the precipitate particles. While such estimates of particle size by x-ray methods depend in part on the method used to assess line widths, they are sufficiently reliable to give at least the order of magnitude of the size. In the present case this proves to be approximately $5\times10^{-6}\,\text{cm}$. Assuming that the particles are spherical and uniformly distributed, it can then easily be calculated that, in a 1% iron alloy, they would be nearly 2×10^{-5} cm apart. This figure varies very little over the range of composition from 1·34% to 0·5%.

Thus we see that in these alloys the x-ray evidence shows the precipitate particles to be of the order of 10^{-5} cm apart. This, as has been shown above, is very close to the optimum distance required for 'locking-in' of dislocations. It is realized that the precipitate, in its early stages, is very likely not spherical but more probably plate-like and that there must be a considerable range in particle size. Nevertheless, the approximate agreement between the deduced distance *d* and that calculated from Orowan's model leads us to believe that a 'locking-in' process of this type occurs and is primarily responsible for the rapid increase observed in resistivity of the precipitated alloys when they are cold worked.

3.4. *Resistivity of Cold Worked Alloys*

It remains for us to consider the variation of resistivity with iron content in the case of the cold worked alloys. Typical curves for a number of degrees of cold working have already been published (see Hetherington and Reekie 1951). A significant feature of these results which will be at once remarked is the similarity in shape of all the curves to that for the annealed alloys, even after cold working has been carried out to 95% reduction of area. At first it might be unexpected that the sharp peak observed in the resistivity curves should remain undisplaced for all degress of cold working. This, however, is readily explained on the basis of the mechanism suggested above.

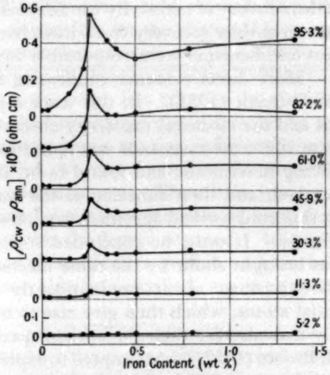


Fig. 7. Additional resistance resulting from cold work as a function of iron content, for the degrees of cold working shown.

Over the incoherent precipitate range a given degree of cold working will result in the generation and subsequent locking-in of about the same number of dislocation lines in *each* alloy, since, as the x-ray evidence shows, the distance between precipitate particles does not vary appreciably over the range from 0·45 to 1·34% iron. Hence $\Delta\rho$, the difference between the resistivities in the cold worked and annealed states, should be approximately constant over this range when plotted as a function of iron content. With the exception of the most heavily cold worked specimens, fig. 7 shows that this is in fact very nearly the case.

Between 0·45 and 0·26% iron, where we believe the precipitate to be partially coherent, a more complicated state of affairs must exist. The sharp rise in $\Delta\rho$ between 0·45 and 0·26% iron indicates that, although coherency strains are probably being relieved by the cold working, locking-in of dislocations or vacancies still takes place, even more effectively than in the incoherent precipitate region. While several possible explanations for this effect suggest themselves, the most likely appears to be that the effective distance between precipitate particles is more nearly the optimum for locking than in the incoherent range.

The fact that cold working has a *positive* effect *below* 0·25%, even though precipitation and consequent depletion of solute atoms is being induced, is again readily explicable in terms of locking-in of the generated vacancies and dislocations about the precipitated particles. As the iron content decreases to low values there must be a rapidly diminishing amount of precipitate, even for high degrees of cold working, and $\Delta\rho$ consequently falls rapidly to a small value.

§ 4. DISCUSSION

In our consideration of these results so far we have not attempted to decide whether the increase in electrical resistivity consequent on cold working is due to dislocations, to vacancies, to interstitial atoms, or to other types of lattice disturbance. In all probability several types of irregularities are effective in varying degrees.

There is a good deal of evidence now available to suggest that dislocations are not alone responsible for the resistivity increase. Molenaar and Aarts (1950) showed that appreciable recovery of resistivity can occur in cold worked materials at much lower temperatures than recovery from work hardening, and concluded therefore that different mechanisms were responsible for increase of resistance and work hardening. More direct evidence confirming this hypothesis is given by the recent work of Dugdale (1952). In this work a platinum specimen was irradiated by neutrons and the electrical resistivity observed to increase by about 0·5%. On annealing at 100°c the resistance was found to recover, the activation energy for recovery being determined and found to be about 1·2 ev. A similar specimen was cold worked, and then annealed at the same temperature. The resistance recovery in this cold worked specimen was observed to have the same value of activation energy. It must be concluded from these results that the resistivity increase was brought about by the same mechanism in the two cases. Since neutron irradiation must result predominantly in the production of vacancies and interstitial atoms, which then give rise to increased resistivity, the same is most probably true also for the cold worked specimen in this case.

On the other hand, Broom (1952) has attempted to explain both work hardening and resistivity increase in terms of dislocations. He assumes simply that various features of dislocation arrays, such as their shape, size, and geometrical arrangement, influence hardness and resistivity in different ways. A decrease in size, for example, of the faults resulting from dislocations would be expected to have much greater influence in recovery of resistance than in softening of a work hardened material.

In studies on the saturation magnetization curves of heavily cold worked polycrystalline materials Brown (1940) has noted that the results can be explained on the assumption of the existence of a *linear* imperfection in the material. This would imply that, in *heavily* worked materials at any rate, there is a relatively high density of dislocations (linear imperfections) as opposed to vacancies (point imperfections). As Seitz (1952) observes, this not necessarily at variance with the supposition that vacancies are mainly responsible for resistance changes in the earlier stages of cold work. In this connection we also recall that the neutron irradiation experiments of Dugdale (1952) gave a resistivity increase, attributed to vacancies, amounting only to 0·5%; a relatively small amount of cold working would result in a resistivity change of this amount, as is evident from fig. 4. Hence it appears possible that when a material is very heavily cold worked the distortional effects due to dislocations may become appreciably more effective in causing resistivity increases than those arising from vacancies. The present results are not inconsistent with this view, and the resistivity rise of the alloys of higher iron content, at high degrees of cold working, is very probably due mainly to dislocation loops being locked around the precipitate in the manner already discussed. For *smaller* amounts of cold work on the other hand, vacancies, generated from moving dislocations in one of the several ways envisaged by Seitz (1952) and by Mott (1952), may be responsible for most of the resistivity increase in this region.

A rough estimate can be obtained of the number of vacancies which would be necessary to produce the observed resistivity change at any degree of cold work if we make an assumption regarding the effect of a vacancy on the resistance of the matrix. Linde (1950) has found that the resistivity of copper increases very nearly linearly as substitutional alloying solute atoms are added. The increase varies from zero to about 10^{-5} ohm cm at most. The observed change in resistivity of the 0·5% alloy when cold worked 95% is $0{\cdot}313 \times 10^{-6}$ ohm cm. If vacancies alone are responsible for the change this would imply that there was 0·0313 atomic % of vacancies present, or $2{\cdot}7 \times 10^{19}$ vacancies per cm^3. This would result in a decrease in density of about 0·003 g cm^{-3} for this amount of cold working. If vacancies were less effective in changing the resistance than has been assumed the density change would have to be proportionately greater. While our density measurements are not sufficiently sensitive to draw positive conclusions on this point, we can state that the observed decrease in density was of the order calculated above. Thus it is just possible for vacancies to account for the observed resistivity increase at high degrees of cold working, provided they are effective to the degree assumed. It is very questionable whether in actual fact a vacancy is as effective as this, and we believe that dislocations, locked-in in the manner already described, are the predominant factor at high deformations.

It is evident that a great deal of information on the relative importance of vacancies and dislocations in increasing the electrical resistivity, and their mode of interaction with the precipitate in these alloys, could be obtained from a study of isothermal recovery curves. It is intended to carry out such observations on heavily cold worked alloys, after deformation at various temperatures.

Acknowledgments

The copper–iron alloys used were very kindly prepared and supplied by the Metals Division of Imperial Chemical Industries Ltd. Financial support for part of this work has been provided by the Defence Research Board of Canada and this is gratefully acknowledged.

REFERENCES

BAUER, O., and HANSEN, M., 1934, *Z. Metallk.*, **26,** 121.

BROOM, T., 1952, *Proc. Phys. Soc.* B, **65,** 871.

BROWN, W. F., 1940, *Phys. Rev.*, **58,** 730.

DUGDALE, R. A., 1952, *Phil. Mag.*, **43,** 912.

GEISLER, A. H., 1951, *Phase Transformations in Solids* (New York : Wiley), p. 387.

HETHERINGTON, F. E., and REEKIE, J., 1951, *J. Appl. Phys.*, **22,** 1293.

HUTCHISON, T. S., and REEKIE, J., 1951, *Phys. Rev.*, **83,** 854.

KAPITZA, P., and MILNER, C. J., 1937, *J. Sci. Instrum.*, **14,** 165.

LINDE, J. O., 1950, *Ann. Phys., Lpz.*, **15,** 219.

MOLENAAR, J., and AARTS, W. H., 1950, *Nature, Lond.*, **166,** 690.

MOTT, N. F., 1952, *Phil. Mag.*, **43,** 1151.

NABARRO, F. R. N., 1948, *Symposium on Internal Stresses in Metals and Alloys* (London : Institute of Metals), p. 248.

OROWAN, E., 1948, *Symposium on Internal Stresses in Metals and Alloys* (London : Institute of Metals), p. 452.

RUTTER, J. W., and REEKIE, J., 1950, *Phys. Rev.*, **78,** 70.

SEITZ, F., 1952, *Advances in Physics* (*Phil. Mag.* Supplement), **1,** 43.

SMITHELLS, C. J., 1949, *Metals Reference Book* (London : Butterworths Scientific Publications), p. 396.

LETTERS TO THE EDITOR

Plastic Deformation and the Electric Strength of Alkali Halide Crystals

Recent investigations (Alger and von Hippel 1949, Calderwood *et al.* 1953) have suggested that the electric strength of alkali halide crystals may be influenced by a number of secondary factors such as cathode material, wave form of applied voltage and the previous mechanical treatment of the specimen. Whilst there is some disagreement about the importance of the first two factors there is little doubt about the effect of mechanical treatment and it has been shown (Calderwood *et al.* 1953) that plastically deformed crystals possess appreciably higher electric strengths than do undeformed crystals.

The electric strengths of the alkali halides are about 10^6 v cm^{-1} and therefore mechanical pressures in excess of about 1 kg cm^{-2} are developed across the specimen, before breakdown occurs, by the electrostatic forces between the electrodes. The shear stress required to cause plastic deformation depends, obviously, on the alkali halide in question. It also depends, for any crystal, on its previous history but for well annealed sodium chloride crystals (Pratt 1952) the necessary resolved shear stress in the slip planes (i.e. (110) planes) may be as low as 1 kg cm^{-2}.

It is therefore of interest to know: (i) whether the mechanical pressure developed by the electrostatic forces between the electrodes is sufficient to deform the crystal before the specimen breaks down, (ii) whether values of electric strength can be obtained for undeformed alkali halide crystals. Only these values are suitable for comparison with theories.

We have investigated these problems, in the case of potassium chloride crystals, by the following experiment. The photo-elastic method was used to detect birefringence due to plastic deformation and the apparatus used is illustrated by figure 1. Transparent electrodes, consisting of a solution of KCl in glycerine, were applied to the specimens and contact was made to the voltage source by brass spheres through which holes had been drilled to allow the passage of polarized light. The experimental procedure consisted of applying standard 1 : 50 μsec voltage impulses of successively increasing amplitude, whilst simultaneously the region of maximum electric stress, at the base of the recess in the specimen, was observed through the microscope.

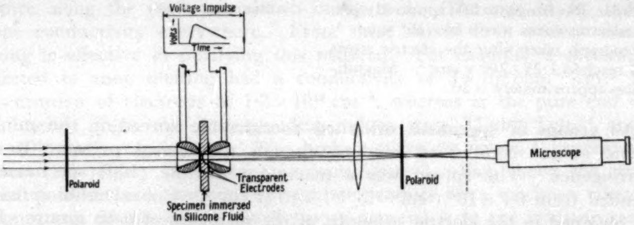


Figure 1. Schematic diagram of apparatus for detecting stress birefringence.

Observations were made at room temperature on three groups of specimens. The electric stress was applied in the (100) crystallographic direction to the specimens of one group, in the (110) direction to those of another group and in the (111) direction to the specimens of the final group.

Before the voltage impulses were applied the specimens exhibited no stress birefringence. Stress birefringence was observed, however, before breakdown occurred in all of the (111) and the (110) specimens. It first appeared when the electric stress exceeded about $0{\cdot}8 \times 10^{6}$ v cm^{-1}. With about 20% of the (100) specimens breakdown occurred without the previous appearance of birefringence. Each of these specimens possessed a low value of electric strength between about $0{\cdot}6 \times 10^{6}$ v cm^{-1} and $0{\cdot}7 \times 10^{6}$ v cm^{-1}. In figure 2 the stress birefringence observed in a (111) specimen is illustrated.

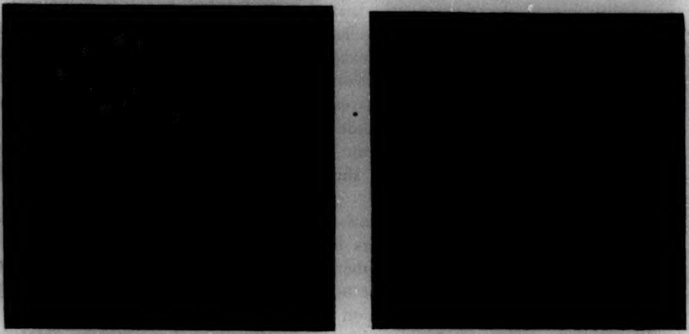


Figure 2. Illustrating stress pattern caused by mechanical pressure developed between electrodes during application of the voltage.

(*a*) Stress pattern in a (111) specimen. (111) crystallographic direction $\perp$ to plane of paper. Birefringence first appeared when the electric stress was $0{\cdot}85 \times 10^{6}$ v cm^{-1}. Photograph taken after the electric stress had reached $1{\cdot}25 \times 10^{6}$ v cm^{-1}. Magnification approximately $\times 30$.

(*b*) Illustrates appearance of a typical specimen viewed before application of voltage impulses.

All groups of specimens exhibited considerable spread in the minimum electric stress required to produce enough mechanical pressure to cause birefringence. The spread was a maximum for the (100) specimens and extended from $0{\cdot}7 \times 10^{6}$ v cm^{-1} to $1{\cdot}1 \times 10^{6}$ v cm^{-1}. Considerable spread was also obtained in the electric strengths of the members of each group. It was observed that the highest values of electric strength were obtained from those specimens allowing the most transmission of light before breakdown occurred, i.e. the most deformed specimens possessed the highest electric strengths.

One definite conclusion may be drawn from the experiment. If values of electric strength are required for comparison with theoretical predictions considerable error is likely to result from selecting the maximum value as the most reliable of a group of comparable measurements. Unfortunately the values generally quoted in the literature appear to have been obtained in this way. The question still remains, what are the values of electric strengths for undeformed alkali halide crystals ? The fact that a number of (100) specimens were observed to break down without previously exhibiting stress birefringence possibly has some bearing on this. There appear to be two possible explanations of this phenomenon. One, which we believe to be most unlikely, is that the specimens were faulty. The other is that they possessed the most ordered crystal structures of the group and consequently their electric strengths were lower than average and also less than the electric stress required to produce sufficient mechanical pressure to cause slip. If the former possibility is excluded a value of electric strength between $0{\cdot}6 \times 10^{6}\ \mathrm{v\ cm^{-1}}$ and $0{\cdot}7 \times 10^{6}\ \mathrm{v\ cm^{-1}}$ corresponds to undeformed crystals of potassium chloride and is to be compared with the value of $1{\cdot}0 \times 10^{6}\ \mathrm{v\ cm^{-1}}$ which is usually quoted (Whitehead 1951) for this material at room temperature.

Department of Electrical Engineering,
University of Liverpool.
9th September 1953.

R. Cooper.
A. A. Wallace.

Alger, R. S., and von Hippel, A., 1949, *Phys. Rev.*, **76**, 127.
Calderwood, J. H., Cooper, R., and Wallace, A. A., 1953, *Proc. Instn. Elect. Engrs.*, **100**, Pt. IIa, no. 3, 105.
Pratt, P. L., 1952, *Atomic Energy Research Establishment Report* M/R 833.
Whitehead, S., *Dielectric Breakdown of Solids* (Oxford : University Press).

A Note on the Semiconducting Compound InSb

The compound InSb has been prepared by direct fusion of the elements in evacuated silica vessels, using Johnson, Matthey 'Specpure' materials. The ingot was then purified by zone melting under vacuum (to an extent equivalent to ten passes of a single zone) and annealed at 350°c, followed by slow cooling to room temperature over a period of 48 hours. The material thus prepared was polycrystalline with about 15 grains per cm^{2} of an etched face.

The conductivity σ and Hall constant R_{H} of the material as a function of distance along the ingot are shown in figure 1. The sign of R_{H} indicated n-type conductivity everywhere. From these curves it is evident that zone melting is effective in purifying this material. For example, a specimen not subjected to zone melting had a conductivity of $3 \times 10^{3}\ \mathrm{ohm^{-1}\ cm^{-1}}$ and a concentration of electrons of $1{\cdot}2 \times 10^{18}\ \mathrm{cm^{-3}}$, whereas at the pure end of the zone melted ingot the corresponding values were $33\ \mathrm{ohm^{-1}\ cm^{-1}}$ and less than $10^{16}\ \mathrm{cm^{-3}}$ respectively. The broken curve in figure 1 represents the product $(8/3\pi)R_{\mathrm{H}}\sigma$. It will be seen from figure 2, which shows the variation of Hall constant and conductivity with temperature for a specimen taken from the pure end of the ingot, that the pure material is in the intrinsic range at room temperature. It follows, therefore, that the product quoted above gives only the difference of the electron and hole mobilities. However, since the product reaches a value of $41\,000\ \mathrm{cm^{2}\ v^{-1}\ sec^{-1}}$ at the centre of the ingot, it is evident that the electron mobility in this material is higher than any other values previously recorded (Breckenridge 1953, Pearson and Tanenbaum 1953, Welker 1952). It is possible that the increase in mobility can be attributed to the annealing treatment since we have found that this process has a significant influence on the mobility. For example, in one specimen $(8/3\pi)R_{\mathrm{H}}\sigma$ increased by a factor of 6 on annealing.

In figure 2 conductivity and Hall constant as a function of temperature are plotted in the form $\log_{10}(\sigma T^{-3/2})$ against $1/T$ (neglecting change of mobility with temperature) and $\log_{10}(R_H T^{3/2})$ against $1/T$. The intrinsic activation energy deduced from the straight part of the latter curve is 0·38 ev. Previous values quoted are 0·40 ev (Breckenridge 1953) and 0·53 ev (Welker 1952). The curve $(8/3\pi)R_H\sigma$ for this specimen is also plotted in figure 2, from which it can be seen that the fall in mobility towards high temperatures above 40°c is smaller than would be expected from the $T^{-3/2}$ law for lattice scattering. The decrease in R_H and the product $(8/3\pi)R_H\sigma$ at the lowest temperatures of the measurements indicates the presence of p-type impurities in the specimen. This is supported by the apparent fall in mobility at the purer end of the ingot as seen in figure 1, which could be explained by the segregation of some p-type impurities towards this end.

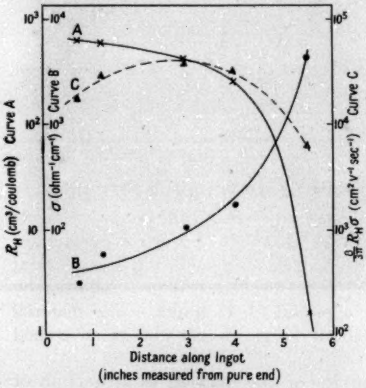


Figure 1.

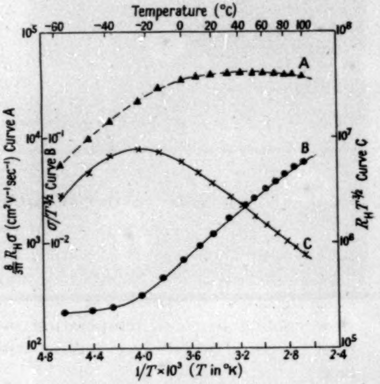


Figure 2.

Optical transmission measurements on bulk specimens have confirmed the presence of an absorption edge corresponding approximately to the above energy gap. In addition, an absorption tail has been observed extending to as far as 9 μ. This is presumably due to the presence of either foreign atoms or defects of thermal origin.

The authors wish to make acknowledgment to Mr. P. Benjamin for his assistance in some of the measurements and to the Admiralty for permission to publish this note.

Note added in proof. Since this letter was written, articles on InSb by M. Tanenbaum and J. P. Maita (*Phys. Rev.*, 1953, **91**, 1009) and by H. Weiss (*Z. Naturforsch.*, 1953, **8a**, 463) have appeared.

Services Electronics Research Laboratory,
Baldock, Herts.
1st September 1953.

F. A. Cunnell.
E. W. Saker.
J. T. Edmond.

Breckenridge, R. G., 1953, *Phys. Rev.*, **90**, 488.
Pearson, G. L., and Tanenbaum, M., 1953, *Phys. Rev.*, **90**, 153.
Welker, H., 1952, *Z. Naturforsch.*, **11**, 744.

The Dielectric Constant of Amorphous Selenium at Wavelengths of 1 cm and 3 cm

Since the dielectric constant of amorphous selenium at infra-red wavelengths beyond 2 microns is very close to 6·0 (Gebbie and Saker 1951, Dowd 1951) whereas at low radio frequencies the value is 6·3 (Tamman and Boehme 1931, D. H. Clarke, unpublished), the value at intermediate wavelengths is of some interest. Gebbie and Kiely (1952) have made measurements at a wavelength of 3 cm using waveguide techniques and give a value of 5·97 (standard deviation 0·04). In an attempt to add to this knowledge the present writers have made measurements on a number of selenium samples at wavelengths of 3 cm and 1·2 cm using cylindrical cavity resonators operating in the H_{01} mode. A Pound frequency stabilizer was incorporated for the measurements in the 1·2 cm region and was capable of holding the frequency constant to 1 part in 10^5. Highest purity selenium (Johnson, Matthey 'chemically refined') was used in the production of the specimens, and particular care was taken to reduce the amount of crystalline impurity to the lowest possible value. As a general guide to the freedom from crystallinity it was confirmed that all specimens were transparent to near infra-red radiation. The results obtained on eight specimens are as follows:

Specimen	A	B	C	C	D	D	D	E	F	G	H
Thickness (mm)	3·187	3·200	1·588	1·588	1·588	1·588	1·588	3·894	2·946	4·928	3·912
λ (cm)	1·2749	1·2749	1·2823	1·2823	1·2820	1·2821	1·0761	3·0926	3·0382	3·0382	3·0382
κ/κ_0	6·27	6·31	6·37	6·40	6·39	6·35	6·30	6·37	6·37	6·30	6·32

The mean value, when weighted for the reliability of individual specimens as regards freedom from defects, is $6{\cdot}37_6$.

Two specimens which were allowed to partially crystallize by prolonged cooling gave values of 6·73 and 7·09.

There is no obvious reason why the present results differ from those of Gebbie and Kiely, but in general it is expected that measurements made using cavity resonators will be considerably more reliable than those made using waveguides. If the present results are correct it appears that the dielectric constant of amorphous selenium at 1·2 cm and 3 cm is essentially the same as that at low frequency. A change of dielectric constant will therefore take place in the long wave infra-red region, and the associated absorption may affect the usefulness of amorphous selenium as an optical material for infra-red wavelengths (Gebbie and Cannon 1952, Gebbie and Saker 1951).

Thanks are due to Professor Willis Jackson and Mr. D. H. Clarke for their interest. One of us (E. W. S.) wishes to thank the Admiralty for permission to publish this letter.

Imperial College, London S.W.7. Y. KLINGER.
Services Electronics Research Laboratory, E. W. SAKER.
Baldock, Herts.
4th August 1953.

DOWD, J. J., 1951, *Proc. Phys. Soc.* B, **64,** 783.
GEBBIE, H. A., and CANNON, C. G., 1952, *J. Opt. Soc. Amer.*, **42,** 277.
GEBBIE, H. A., and KIELY, D. G., 1952, *Proc. Phys. Soc.* B, **65,** 553.
GEBBIE, H. A., and SAKER, E. W., 1951, *Proc. Phys. Soc.* B, **64,** 360.
TAMMAN, G., and BOEHME, W., 1931, *Z. Anorg. Chem.*, **197,** 1.

The Reduction of Rectifier Noise by Illumination

The hypothesis of current noise generation in germanium proposed by Montgomery (1952) and the increased current gain in germanium transistors at small emitter currents observed by Sittner (1952) suggested to us that the noise in a germanium diode should be reduced when the crystal is illuminated. This effect has been observed experimentally.

The noise in a germanium rectifier was measured at room temperature using an amplifier tuned to 16·5 c/s with a bandwidth of 1·4 c/s. The germanium surface and the point contact could be flooded with light from a 240-watt d.c. tungsten filament lamp, the heat radiation from the lamp being absorbed by a 1-inch slab of Perspex. Different n-type germanium rectifiers were tested and typical results are shown in figures 1 and 2. In figure 1 the noise power per unit dynamic resistance is plotted as a function of reverse bias current in the rectifier. The applied voltages are also marked on the graph. Under dark or illuminated conditions the noise increases in a fairly regular manner with bias current though the increase is more rapid for the illuminated diode. The noise at constant bias current decreases on illumination, the decrease being greater for small bias currents. In figure 2 the decrease of noise power per unit dynamic resistance is plotted against reverse bias current for various separations between the lamp and the germanium surface. The reduction of noise for a given bias current increases as the light intensity increases.

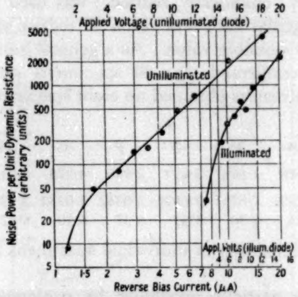


Figure 1. Noise power per unit resistance plotted against reverse bias current for a germanium diode before and during illumination.

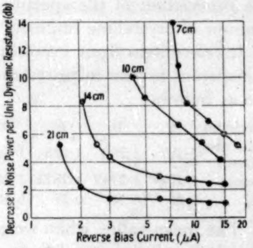


Figure 2. Decrease of noise power per unit resistance plotted against reverse bias current for different separations between the light source and the rectifier.

These results may be partially explained as follows. Montgomery has advanced the hypothesis that noise in germanium filamentary devices is caused by holes and electrons that are generated in the interior and at the surface of the semiconductor. Sittner has shown that the current magnification factor α in a type A germanium transistor increases at very low emitter currents. More recently, Cooke-Yarborough and Stephen (1954) have obtained typical increases of α at room temperature from 2 at large emitter currents (of order 0·1 mA) to 8 at small emitter currents (of order 1 μA). A point contact on germanium biased in the reverse direction collects holes thermally generated near the contact and it is assumed that this hole current fluctuates in such a manner as to give rise to the observed noise spectrum. This current is very small ($\ll 1\,\mu\text{A}$) and the resulting high value of α at the collector produces large current fluctuations in the rectifier and hence a large noise figure. When the crystal is illuminated a large hole current flows to the point contact which is then a collector with a low value of α. The fluctuations in the collector current caused by the thermally generated hole current are then smaller than for the unilluminated crystal and the noise figure is reduced. On illumination, the decrease of noise would be expected to be less than $[\alpha^2(\text{small } I_e)]/[\alpha^2(\text{large } I_e)]$ or 12 db for the figures quoted above. This explanation is also consistent with the results shown in figure 2.

The authors are indebted to the Chief Scientist, Ministry of Supply, and to the Controller, H.M. Stationery Office, for permission to publish this letter.

Radar Research Establishment,
Great Malvern, Worcs.
25th September 1953

J. W. Granville.
A. F. Gibson.

Cooke-Yarborough, E. H., and Stephen, J. H., 1954, *J. Instn. Elect. Engrs.*, in the press.
Montgomery, H. C., 1952, *Bell Syst. Tech. J.*, **31**, 950.
Sittner, W. R., 1952, *Proc. Inst. Radio Engrs., N.Y.*, **40**, 448.

OBITUARY NOTICES

WILLIAM DAMPIER

By the death of Sir William Dampier (formerly William Cecil Dampier Whetham) on 11th December 1952 the Physical Society has lost a senior Fellow and Cambridge one of its oldest and best loved sons, who was recently described by Sir Lionel Whitby as "a most distinguished scientist, agriculturalist and sociologist" (*Cambridge University Reporter*, October 1953). Dampier, son of the late C. L. Whetham, was born in London in 1867, and one of his grandfathers was Lord Mayor of London in 1878. He was educated at Trinity College, Cambridge, and took his B.A. degree in 1889. He held both the Coutts Trotter Studentship and the Clerk Maxwell Scholarship, and in 1891 was elected to a Fellowship which he retained throughout his long life. For many years he held a College lectureship, and gave courses in Heat and Electricity for students working for Part I of the Natural Sciences Tripos. In the earlier years these lectures were delivered in Trinity, but were later transferred to the Cavendish Laboratory, where more accommodation and better facilities for lecture experiments were available. Many of those who attended his lectures in the days before the First World War must still have vivid memories of his habit of pacing slowly backwards and forwards behind the lecture table, his deliberate but admirably clear delivery, the charm and lucidity of his exposition, and his gift for arousing interest in physics as an exciting and rapidly advancing subject. His elegant treatment of the basic principles of thermodynamics is preserved for us in the first chapter of his *Theory of Solution*, published in 1902. In more popular vein was his *Recent Development of Physical Science* (1904), which gave him full scope for the display of his considerable literary gifts, and revealed a sense of humour that prompted him to preface a chapter on Atoms and Aether with the quotation: "Oh, dear! What can the matter be?—*Old Song*." His characteristic pleasing style is to be found even in his textbooks, such as *Theory of Experimental Electricity* (1905), a work that was deservedly popular in its day, but which may have resulted in some loss of freshness and spontaneity in the lectures on which it was based, and so caused some of the later students to complain that what they listened to was "all in the book".

From 1889 onwards Dampier worked in the Cavendish Laboratory under J. J. Thomson, and between 1890 and 1905 he produced a long series of important papers published in the *Philosophical Transactions*, the *Proceedings of the Royal Society* and the *Philosophical Magazine*. Some of his earlier experimental work was concerned with the flow of liquids through tubes, but most of his researches were directed towards the elucidation of the properties of electrolytic ions in solution, a subject which he made his own and on which he wrote authoritatively in his *Theory of Solution*. Part of this work was done in collaboration with E. H. Griffiths. Dampier's eminence in his chosen field led to his election in 1901 to the Fellowship of the Royal Society. After he became Tutor and, later, Senior Tutor of Trinity, an office which he held until 1917, he found that the heavy burden of administrative duties left him little time for original research, but though his experimental work ceased his interest remained, and he began gradually to collect material for his *History of Science*, which was not published until 1929.

Though a Londoner by birth, Dampier was always greatly interested in country life, and this led him to purchase an estate in Devonshire, and to turn his attention to problems of agriculture. Later (1917) he inherited a family estate. During the First World War he worked in the Food Production Department of the Ministry of Agriculture, and from that time onwards he began to be recognized as an authority on agricultural matters and was frequently consulted by the Government. He played a leading part in preparing the ground for the setting up of the Agricultural Research Council, of which he became first secretary, and to which was entrusted the administration of public funds devoted to agricultural research. In 1927 he published his book *Politics and the Land.* In 1931 he was knighted in recognition of his public services, and in the same year he was awarded the Gold Medal of the Royal Agricultural Society, of which he became Vice-President in 1945. He was also a member of the Argicultural Wages Board, and was a Fellow of Winchester College from 1917–47.

In 1897 Dampier married Catherine Holt, daughter of the late R. D. Holt of Liverpool, and they had five daughters. He was a man of quiet charm, kindly nature and friendly disposition, and his success both as College Tutor and also in his public relations was due in no small measure to his deep understanding of human nature. It was characteristic of him that when he had to ask a junior to give a lecture for him he would never offer detailed instructions. "Treat the subject in your own way" he would say "and you will make a success of it". He will be much missed by all who knew him.

G. STEAD.

ROBERT DONALDSON

By the sudden death, on 5th November 1953, of Robert Donaldson, Head of the Colorimetry Section of the Light Division of the National Physical Laboratory since 1933, his many friends and colleagues suffered a great personal loss, and the branch of science which he had made his own was deprived of one of its leading exponents in the fullness of his career.

Donaldson was born on 8th June 1904, in the fishing village of Port-Seton on the Firth of Forth, and was the oldest of three brothers, all of whom were later to gain distinction as students at the University of Edinburgh. His principal studies were Physics and Mathematics—under Professor C. G. Darwin, who was later to become Director of the National Physical Laboratory. After graduation he remained at the University for two years as Demonstrator in the Department of Natural Philosophy.

He joined the staff of N.P.L. on 3rd July 1928, and was first employed as an assistant in the work which culminated in the adoption by the Commission Internationale de L'Eclairage, in 1931, of the first internationally agreed system of colorimetric standardization. It is a curious coincidence that the quantitative basis of that system, having stood the tests of twenty years, is only now being rechecked and extended with the greatly improved resources of modern equipment and technique, and that Donaldson, in collaboration with Dr. W. S. Stiles, was engaged on this work at the time of his death : so in one sense his first job was also his last. But much happened in the intervening years. When he joined us at N.P.L. his interest was more in the mathematical aspects of physics than the experimental; and though, as events showed, he had a natural flair for experiment he had no special leaning to any particular field. However, he quickly mastered the apparent complexities of colorimetry and heterochromatic photometry and became fascinated by the interesting problems, physical, physiological, and mathematical, which he could foresee would arise in the course of their development to meet the ever-increasing requirements of industry. Just when useful investigation in any field of this kind ends depends on the imagination and foresight of those working in it. To Donaldson colorimetry had no end, and except during the war, when he participated in the general optical work of the Light Division, he has been primarily occupied, by his own choice, with this subject and its ramifications, gaining for himself an international reputation.

But he avoided the narrowness often found in the expert. He always retained his interest in mathematics and was a keen follower of all significant developments in relativity and quantum physics, bringing a balanced philosophical judgment to discussions of these matters or others of fundamental scientific import. This width of interest helped his work as a specialist by keeping his feet on the ground and his eyes on essentials. Like most of us he may sometimes have failed to see the most direct solution of some problem of theory or design, but if so the writer knows of no instance.

His many contributions to his subject are too well known to all workers in the field to need recapitulation, especially here where we are more concerned with what we have lost of future promise than with what we have had from him.

He enjoyed the respect and esteem not only of his colleagues in N.P.L., but also of the many with whom he was associated on committees of other organizations. Normally of a cheerful—and cheering—disposition, which was often a source of help and encouragement to others, he nevertheless suffered from shyness which showed itself as a dislike, sometimes amounting to a dread, of personal prominence. For example, he could not be induced to accept nomination for Chairmanship of the Society's Colour Group, of which he has been an enthusiastic member from its inception, though strongly pressed to do so on several occasions. He was loath to give lectures, or to participate in discussions among unfamiliar groups, if he could possibly avoid doing so; but if it could not be avoided, when he had forgotten his audience in the interest of his subject there could be no clearer exponent of whatever he had to say.

He had little interest in outdoor sports of the more intense varieties, such as football or tennis, but was an enthusiastic golfer. He was fond of music, being an efficient pianist and a fair violinist. The writer has memories of many pleasant musical evenings in his home during the pre-war years.

In 1934, he married Miss Kathleen Edwards, who was at that time a colleague on the Laboratory staff. They had one son, Ian, now 12 years of age.

J. GUILD.

LEONARD BELLINGHAM

We record with regret the death of Leonard Bellingham who had been a Member of the original Optical Society since 1918 and continued his Membership with the Physical Society and its Optical and Colour Groups. Mr. Bellingham was a founder of the firm Messrs. Bellingham and Stanley, Ltd., having gained his skill in the making of sensitive apparatus with Messrs. Adam Hilger, Ltd.

EDWARD FELIX HERROUN

We record with regret the death of Edward Felix Herroun who had been a Member of the Physical Society since 1885. Herroun was born in 1862 and was educated at Greenwich and King's College, London. In 1881 he gained the Daniell Scholarship in chemistry. His main interest was in research on magnetites, and this work he continued for many years after his retirement.

W. H. TOWNS

We record with regret the death of W. H. Towns who had been a Member of the original Optical Society since 1918 and continued his Membership with the Physical Society and its Optical Group. He maintained his interest in optics throughout his very long life and died on 3rd September 1953 at the age of 92.

LEWIS FRY RICHARDSON

Lewis Fry Richardson was born in 1881, the youngest of the seven children of David Richardson, who was the owner of an old-established leather works in Newcastle-on-Tyne. All the children of this well-to-do Quaker family showed unusual ability, not least of them Lewis.

He was educated at Bootham School, York, and at the Universities of Durham and Cambridge, where he gained a First Class in the Natural Sciences Tripos in 1903. He was for a few years at the National Physical Laboratory, but his interests gradually changed to geophysics and meteorology: he was for a time Superintendent of Eskdalemuir Geophysical Observatory and was elected a Fellow of the Royal Society in 1926. The last twelve years of Richardson's professional life were spent as Principal of Paisley Technical College.

Richardson had been a Fellow of the Physical Society since 1922, and from 1921 to 1924 was Honorary Secretary of the Royal Meteorological Society. During the First World War he served with the Friends' Ambulance Unit in France. He was a man always popular with his students, and always gave an impression of youth despite his white hair.

REVIEWS OF BOOKS

Logic for Mathematicians, by J. Barkley Rosser. Pp. xiv + 530. (London: McGraw-Hill, 1953.) 85*s*.

It is related of the late Professor Hardy that, when lecturing on the theory of divergent series, he came to a point in a proof where he said " Now, it is obvious that . . . ". He stopped. There was a long, silent pause. Then he said " But is it obvious ? ". After another long pause he left the room. Ten minutes later he came back, said " Yes, it *is* obvious " and went on with the lecture. Another story, the characters in which must remain anonymous, because they are fortunately still with us, is told of the professors of the leading University of the United States of America. It is, that when Professor C says a thing is obvious, it means that the class already saw the point hours ago, and are impatiently waiting for the next step ; when Professor B says a thing is obvious, it is just obvious. When Professor SB says a thing is obvious, if you go away and think about it for a week, you will probably succeed in proving it. But when Professor L says a thing is obvious, you know it is false.

Professor Rosser defines the object of his book as that of giving a formal definition of the word ' obvious ', or, rather, of the phrase ' immediate consequence of ', so that we can tell, by means of a mechanical checking procedure, whether one statement is, or is not, an ' obvious ' consequence of another. He claims no absolute or intrinsic justification for his definition, only that its results seem to coincide with those of the ' judgement of careful mathematicians '. His programme is thus the same as that of Russell and Whitehead, in *Principia Mathematica*, except perhaps that these latter laid claim to an absolute validity which Rosser explicitly disclaims. The present volume is considerably shorter than the three volumes of *Principia Mathematica*, partly because some topics, no longer of great mathematical interest, are omitted, but mainly because Professor Rosser makes use of the powerful results that have been obtained in symbolic logic since 1912 when *Principia* was first published. In fact the present work goes further than did *Principia Mathematica*, in that a detailed discussion is given of the various forms of the axiom of choice, and its modern equivalent, Zorn's lemma. And illustrations of the various forms of argument used are drawn from current mathematical texts, ranging from textbooks of elementary geometry to advanced works on the theory of functions.

In course of his development, from the calculus of statements, through the restricted predicate calculus, to the theory of classes, relations and functions, and finally to the theories of cardinal and of ordinal numbers, Professor Rosser has many illuminating remarks to make about mathematicians' use of symbols. One of the most valuable distinctions to which he calls attention is that between a function f and a function value $f(x)$. The equation (or ' identity ') $x^2-2x+1=(x-1)^2$ is a statement that two functions are equal, while the equation $x^2-2x+1=0$ is a statement that two numbers, one of which is a function value, are equal. Current mathematical notation is unable to distinguish between the function x^2 and the function value x^2, and Rosser shows how Church's notation can be used to overcome this difficulty. Church would write $\lambda x \,.\, x^2$ for the function, reserving x^2 for the function value. The x in $\lambda x \,.\, x^2$ is a ' bound variable ', like the x in $\int_0^1 x^2dx$, and it can be changed to another letter y, for example, without altering the meaning. So that, for example, we have, for all values of x, $(\lambda y \, . \, y^2)(x) = x^2$. There is no doubt in the reviewer's mind that attention to this point of notation alone would save many a student great difficulty in the understanding of the theory of partial differentiation, of Laplace transforms, and especially thermodynamics. Church's notation has long deserved to be better known, and Professor Rosser's book would be worth while if it did no more than bring this about.

But besides helping to do this, and helping to clarify many similar points of difficulty for the aspiring mathematician, Professor Rosser has much to say of mathematical interest to the mature mathematician, aside from his main logical thesis. In particular, his treatment of the theory of sets, cardinals, ordinals and the axiom of choice supersedes the accounts in current books, which are either too old-fashioned, or too brief (as, for example, the modern account given by M. Bourbaki).

The one point of criticism which occurs to the reviewer is that too little attention is given to the outstanding results in metamathematics associated mainly with the name of Gödel. Gödel's theorem about the impossibility of a consistency proof for a formal system of mathematics is mentioned briefly only once, and his main theorem, of which this is a corollary, on the existence of undecidable arithmetical propositions, is not mentioned explicitly at all. Nor is his result about the completeness of the predicate calculus mentioned. Tarski's work on the concept of truth in formal systems is also not mentioned, nor is the Skolem–Löwenheim theorem and its associated paradox concerning the real numbers.

But the author could well answer this criticism by saying that his book does not purport to be a book on metamathematics. A book on this latter subject has just been published by the author's former colleague, S. C. Kleene. And it is certainly true that in his method of presenting symbolic logic, Professor Rosser takes full account of these metamathematical results by explicitly disclaiming finality and completeness for his system. He says :

> " Our symbolic logic is not intended as a model for how mathematicians should think but only as a model of how at the present time they do indeed think. Indeed it is desirable that new and more potent and flexible principles of reasoning be devised and generally accepted, so that distant portions of the mathematical edifice will become more readily accessible. One advantage of a symbolic logic is that it can be made very precise, but an even greater advantage is that it can be changed to fit the circumstances."

Such a forward-looking attitude, welcoming the possibility of change in our modes of thinking, is most refreshing at a time when so many people still cling to the discredited notion that the principles of mathematical thinking are absolute, and that the ' inductive ' and other modes of thinking used in physics and elsewhere are in some sense ' inferior ' to mathematical deduction in its ' immutable glory '.

This book should be read by everyone concerned with the teaching of mathematics, and by every mathematics undergraduate at some stage in his career. It ought also to be in the library of any grammar school, where it can be read by the mathematics master and dipped into by the mathematics scholarship candidates. Anyone who does read it, with some knowledge of first-year University mathematics, can be guaranteed a feast of entertainment as well as instruction.

G. A. BARNARD.

Physical Biochemistry, 2nd Edn., by HENRY B. BULL. Pp. ix + 355. (New York : Chapman and Hall, 1951.) 46*s*.

The author has attempted to present a course of physical chemistry to students of the biological sciences, covering in some 350 pages all relevant branches of general theoretical and physical chemistry as well as many specialized topics of importance in the investigation of biological systems. In general the chapters begin at a very elementary level, suggesting that little or no previous knowledge is required of the reader; at the same time each subject is brought to a stage of development at which its application to biological problems can be discussed at some length. In spite of miracles of compression, this unavoidably leads to unsatisfactory features, and the treatment of fundamental physical chemistry is hurried, often trivial and sometimes unsound. While the book is not to be recommended as a text for instruction in physical chemistry, the later chapters are written with some authority and collect together a wide field of applications of physical chemistry : they form interesting reading for physical chemists who, like the reviewer, tend to become too narrow in their interests.

D. J. G. IVES.

The Physical Chemistry of Surface Films, by W. D. HARKINS. Pp. xvi + 413. (New York : Reinhold; London : Chapman and Hall, 1952.) 80*s*.

This is a book written by a physical chemist for physical chemists. It is not out of place, however, that it should be reviewed in the *Proceedings of the Physical Society*, because the various phenomena associated with surfaces and surface films are of considerable interest to physicists. Problems requiring a knowledge of the physics and chemistry of surfaces constantly arise both in the research laboratory and in industry. The number of modern industrial processes which depend, for example, upon the spreading of liquids and the wetting of solids is surprisingly large and results obtained in the laboratory are frequently required for application on the pilot plant scale. One hesitates before referring here to the facetious generalization that physicists do accurate work with impure substances whereas chemists make inaccurate measurements on pure substances. This book will convince any reader that careful attention to both accuracy in measurement and to the state of the material under investigation is essential to the study of surface behaviour.

The author, the late Professor W. D. Harkins, completed the work of writing this book just before his sudden death in March 1951. A perusal of the list of papers—270 in all—published by him either alone or in collaboration with others between the years 1907 and 1950 shows that it is not surprising that almost every page of the book contains references to methods devised, measurements made and theories proposed by him and his fellow workers. The responsibility of obtaining accurate experimental results to test his theories was accepted by Harkins as a serious and personal matter. Uncritical acceptance of published data was not part of his plan. He did the work himself, and as a result was able to bring some order to a field in which unreliable measurements have been the cause of much misunderstanding.

We do not have here a textbook set out in the orderly step by step fashion which would be appreciated by a student on first acquaintance with the subject, It is not a classical exposition of a single connected topic from which all the corners have been smoothed away. It is on the other hand a vigorous account of work done and research successfully carried out in a workmanlike effort to understand and relate phenomena in an important field. Apart from being a general source of information on the subject, it is valuable as an account of the work of one man and his school.

After an introduction by T. F. Young there are six chapters dealing respectively with the nature and energetics of surfaces, films on liquids, films on solids, properties of soap solutions, the mechanism of emulsion polymerization, and the role of the electric double layer in the behaviour of lyophobic colloids. The first three are based on articles previously published in *Colloid Chemistry*, edited by J. Alexander, and the last, which has also appeared in the same work, is contributed by E. J. W. Verwey. The first chapter contains a discussion of surface tension and surface energy, and of arguments based on thermodynamics and on considerations of the field in the neighbourhood of atoms, which are used to describe the phenomena occurring at interfaces. There are several sections on the orientation of molecules in surfaces. In this chapter, as indeed throughout the book, there are valuable accounts of carefully conducted experiments together with critical discussions of the errors to be expected and avoided. To take the measurement of surface tension as an example, the drop weight method (developed to a method of high precision by Harkins and Brown) and the capillary height method are discussed in detail. The considerations outlined here are of fundamental importance to those whose work involves the measurement of surface tension. The next chapter, of a hundred pages, deals with the properties of films on liquids, with the spreading of one liquid upon another, with monomolecular films and also with thicker (duplex) films. There is a comprehensive account of experimentally observed phenomena and theories of spreading advanced, for example by Rayleigh, Langmuir, Harkins and others, to explain them. The development in recent years of experimental techniques based on the film balance (which was used by Agnes Pockels and Lord Rayleigh towards the end of the last century) has yielded important results in investigations of films on liquid surfaces. The effect of liquid or adsorbed films on the energy relations of interfaces at solid surfaces are considered in a succeeding chapter on the basis of the idea that the general principles are the same as in the case of liquids except for effects which are due to the crystal structure of the solid substrate. The methods of investigation are necessarily very different. The properties of the ionic aggregates known as micelles play an important part in the chapter on soap solutions and also in that on emulsion polymerization. The useful chapter on the rôle of the electric double layer in the behaviour of lyophobic colloids completes the book.

The book contains a great deal of information and a large number of references to original papers. It will be of value to physicists as well as chemists, and particularly to those physicists investigating surface phenomena in collaboration with colleagues who are chemists. M. R. HOPKINS.

CONTENTS OF SECTION A

Mr. G. R. Satchler. Angular Distribution of γ-Radiation Following a Deuteron Stripping Reaction 1081

Dr. P. Swinbank and Dr. J. Walker. The Decay Scheme of Krypton-83 . 1093

Dr. B. H. Bransden, Dr. A. Dalgarno and Mr. N. M. King. The Application of Variational Methods to Scattering by Ions: II—The Distorted Wave Approximation and the 1s–2s Excitation of Helium Atoms by Electron Impact 1097

Dr. N. H. March and Dr. B. Donovan. Note on the Spin Paramagnetism of a Free Electron Gas 1104

Dr. C. B. Sharma. Absorption Bands of SbSe and SbTe in the Quartz Ultra-Violet Region 3650 to 2200 Å 1109

Prof. D. R. Bates, Mr. U. Opik and Mr. G. Poots. Properties of the Hydrogen Molecular Ion: II—Photo-Ionization from the $1s\sigma_g$, $2s\sigma_g$ and $3s\sigma_g$ States . 1113

Prof. D. R. Bates, Mr. R. T. S. Darling, Mr. S. C. Hawe and Dr. A. L. Stewart. Properties of the Hydrogen Molecular Ion: III—Oscillator Strengths of the $1s\sigma_g$–$2p\pi_u$, $2p\sigma_u$–$3d\pi_g$ and $2p\pi_u$–$3d\pi_g$ Transitions 1124

Prof. H. S. Green and Dr. E. Wolf. A Scalar Representation of Electromagnetic Fields 1129

Dr. P. K. Carroll and Dr. N. D. Sayers. The Band Spectrum of Nitrogen: New Studies of the Triplet Systems 1138

Mr. P. A. Fraser and Mr. W. R. Jarmain. Vibrational Transition Probabilities of Diatomic Molecules: I 1145

Mr. W. R. Jarmain and Mr. P. A. Fraser. Vibrational Transition Probabilities of Diatomic Molecules: II 1153

Dr. R. J. Blin-Stoyle. The Magnetic Moments of Spin $\frac{1}{2}$ Nuclei . . . 1158

Dr. G. G. Hall. The Electronic Structure of some Body-Centred Cubic Metals 1162

Mr. A. Ashmore and Dr. A. V. Crewe. The Multiple Scattering of 7·5 Mev Deuterons in Metals 1172

Research Notes:

Mr. G. A. Jones and Dr. D. H. Wilkinson. The Reaction $^{13}\mathrm{C}(\alpha n)^{16}\mathrm{O}$. 1176

Mr. J. M. Reid and Dr. K. G. McNeill. The Production of a Short-Lived Activity by Irradiation of Pb with X-Rays 1179

Mr. G. Poots. The Effect of Variation of the Dipole Moment with Internuclear Separation on the Relative Intensities of the Second Negative Band System of Oxygen 1181

Letters to the Editor:

Dr. J. G. Valatin. Predissociations and CO Dissociation Energy . . 1185

Dr. E. R. Andrew and Mr. D. Hyndman. Proton Magnetic Resonance Evidence for the Planar Structure of the Urea Molecule 1187

Dr. A. B. Lidiard. Antiferromagnetism in Metals 1188

Obituary Notices:

William Dampier 1191

Robert Donaldson 1192

Leonard Bellingham 1193

Edward Felix Herroun 1194

W. H. Towns 1194

Lewis Fry Richardson 1194

Reviews of Books 1195

Contents of Section B 1200

Subject Index, Section A, Vol. **66** 1201

Index of Authors (with Titles), Section A, Vol. **66** 1211

Index to Reviews of Books, Section A, Vol. **66** 1219

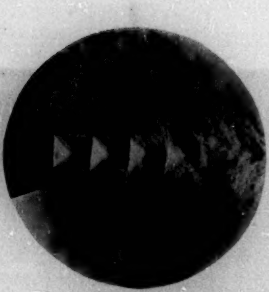


Fig. 6.

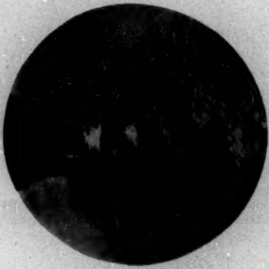


Fig. 7.

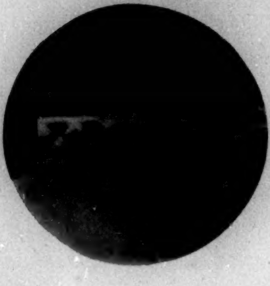


Fig. 8.

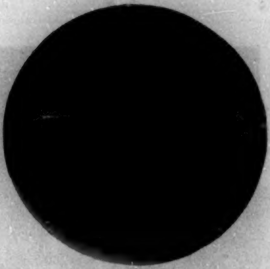


Fig. 9.

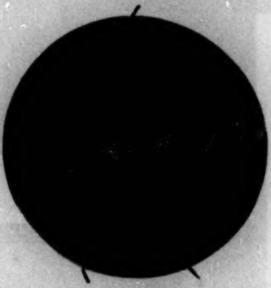


Fig. 10.

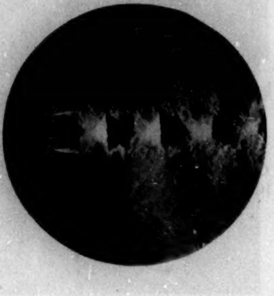


Fig. 14.

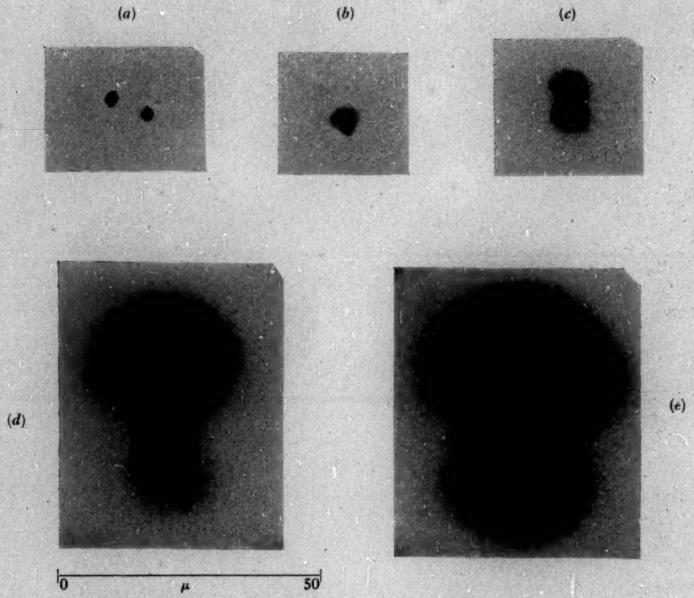


Fig. 2. Microphotographs of the glows produced in phosphors and their mirror images by a fine electron beam; electrons travelling from bottom to top.

(*a*) $CdWO_4$, 10 kv, 120 sec; (*b*) KI, 10 kv, 20 sec; (*c*) KI, 40 kv, 20 sec; (*d*) plastic, 40 kv, 10 sec; (*e*) plastic, 40 kv, 80 sec.

$CdWO_4$ is birefringent: two images are therefore obtained.

PROC. PHYS. SOC. VOL. 66, PT. 12—B (G. N. LANCE AND R. L. PERRY)

PLATE I.

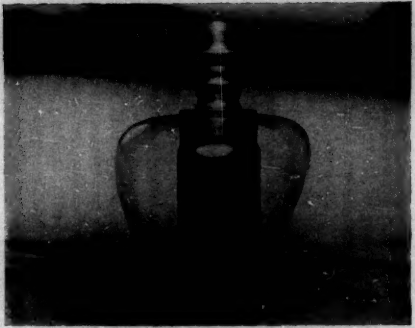


PLATE II.

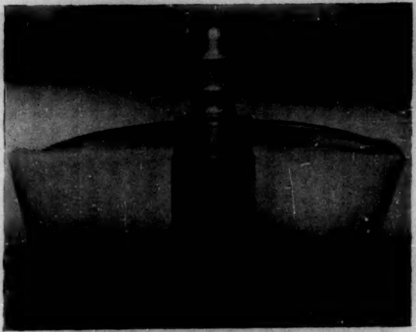


PLATE III.

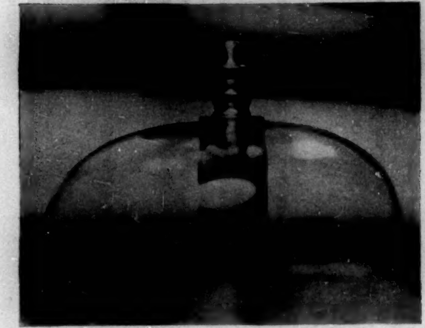
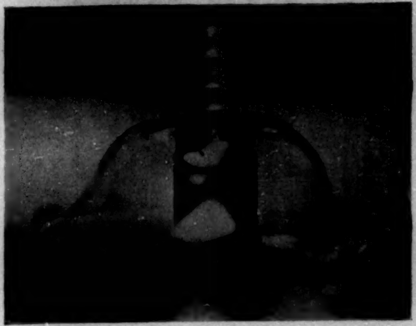


PLATE IV.



PROCEEDINGS OF THE PHYSICAL SOCIETY

SECTION B, 1953—VOL. 66

SUBJECT INDEX

PAGE

Absorption, electromagnetic, in polar liquids, at 3×10^9 c/s by measurement, standing wave method 175
Absorption, injected, in germanium 588
Absorption, monochromatic, in an exponential atmosphere, grazing-incidence integral Ch (x, χ) for (R) 710
Absorption spectra, infra-red, determination from reflection measurements . . 969
Absorption spectra of lead sulphide at different temperatures 194
Additivity of colour equations 548
Afterglow, decay of long period, of alkali halides under cathode-ray excitation . 371
Alcohol and pentane, direct measurement of specific heat at constant volume (R) . 421
Alkali halides, decay of long period afterglow under cathode-ray excitation . . 371
Alloys, iron–nickel, order–disorder transformation in region $FeNi_3$ . . . 221
Aluminium, electron currents in thin oxide films on 533
Aluminium films, thin, reflectivity, and their use in interferometry . . . 963
Aluminium, ionic current and film growth of thin oxide layers on . . . 317
Analogue computer employing principle of Kelvin bridge 185
Anisotropic bodies, electrostatic and steady-current problems 557
Arc discharge, pulsed, experiment on radial motion of charged particles from . 1087
Arc, high pressure, energy equation for positive columns of 169
Asymmetrical temperature gradients in Hg, thermoelectric effects . . . 649
Atmosphere, exponential, grazing-incidence integral Ch(x, χ) for monochromatic absorption in (R) 710

Barkhausen effect in single crystals 633
Beta-ray energies, absolute determination of, development of new method . . 201
Beta-ray spectrometer, *see* Spectrometer, β-ray.
Birefringence, stress, mechanism in amorphous solids. 617
Breakdown and electrical conduction in liquid dielectrics . . . 85, corr. 815
Breakdown, electrical, of gases in non-uniform fields at low pressure . . . 345
Brittleness, room temperature, of chromium (L) 515
Bubble jets, sensitive (L) 258

Caesium, ^{137}Cs γ-line, precision measurement 54
Calomel, formed on liquid mercury by anodic polarization, structure and orientation 129
Capacitor method, for measurement of contact potential difference in high vacuum 493
Carrier injection and extraction in lead sulphide 50
Cathode material, dependence of dielectric strength of pure liquids on (R) . . 425
Cathode material, influence on electric strength of potassium bromide crystals (L) . 716
Cathode-ray excitation, *see* Excitation, cathode-ray.
Charged particles, radial motion of, from pulsed arc discharge, experiment on . 1087
Chromium, room temperature brittleness (L) 515
Coefficient, distribution, *see* Distribution coefficient.
Collisional processes and similarity in high-frequency discharges in helium . . 17
Colour equations, additivity 548
Computing problem, microwave (R) 987
Conduction, electrical, and breakdown in liquid dielectrics . . . . 85, corr. 815
Conduction, *see also* Low temperature conduction.
Conductivity, electrical, *see* Electrical conductivity.
Contact noise in semiconductors (L) 334
Contact potential difference in high vacuum, simple varying capacitor method for measurement 493
Contacts, area, injecting, on Ge, forward characteristic 841
Contacts, emitter, carrier injecting properties at normal and moderately elevated temperatures 845

PAGE
Contacts, point, to semiconductors, electro-thermal behaviour 570
Contacts, point, injecting, forward characteristic, theory 833
Contacts, point, Ge rectifier, noise spectra measurements 1017
Contacts, point, *see also under* Diodes.
Convection, free, over parallel sources of heat 393
Convergence, role in stereoscopic vision (L) 513
' Corona pressure ' and negative Joshi effect (L) 256
Creep, transient, in pure metals 459
Crystallization and distribution coefficient of indium in germanium (L) 60
Crystal, single, disc specimens of silicon–iron, ferromagnetic domain processes in (R) 712
Crystal structures of gutta percha 7
Crystals, alkali halide, plastic deformation and electric strength (L) 1113
Crystals, iron, surface deformation by unidirectional abrasion (L) 717
Crystals, potassium bromide, influence of cathode material on electric strength (L) 716
Crystals, KCl and NaCl, variation of electric strength with temperature 74
Crystals, single, Barkhausen effect 633
Crystals, single, germanium, optical constants (L) 512
Crystals, single, of In and Sn, measurement of thermal expansion 887
Crystals, single, of MgO at high temperatures, electrical conductivity of 281
Crystals, single metal, heat flow and growth from melt 525
Crystals, single, of PbS, photoelectromagnetic and photoconductive effects in 993
Crystals, single, of tin, effect of added metallic impurity and closed moulds on growth from melt 65
Crystals, stearic acid, interferometric studies of growth, and their optical properties 414, corr. 523
Crystals, stearic acid, optical properties and growth of, further note (L) 989
Cuprous oxide rectifier characteristics 622
Current noise and electrical conductivity of carbon resistors 680
Current, *see also* Electron current.

Decay of radio echoes from meteor trails (L) 150
Deflector, magnetic, *see* Magnetic deflector.
Deformation, plastic, and electric strength of alkali halide crystals (L) 1113
Dielectric constant of amorphous Se at wavelengths of 1 cm and 3 cm (L) 1117
Dielectric constant of liquid containing spherical particles (L) 611
Dielectric, liquid, electrical conduction and breakdown in 85, corr. 815
Dielectric strength of pure liquids, dependence on cathode material (R) 425
Diffraction images in systems with an annular aperture 145
Diffraction patterns, *see* X-ray diffraction patterns.
Diffusion theory of rectifier barriers, influence of mobility variation in high fields (L) 430
Dilatometer, photoelectric recording, for measurement of thermal expansion of In and Sn single crystals 887
Diodes, point contact, thermal effects at 826
Discharge, low-frequency electrodeless, secondary processes (L) 431
Discharges, caesium, in magnetic field, spectroscopic study (L) 898
Discharges, caesium, spectroscopic study in magnetic field 33
Discharges, *see also* Electrical discharges, High-frequency discharges.
Dislocation nodes in face-centred cubic lattices 481
Dislocations, array of, mean shear stress, and latent hardening 2
Dislocations, edge, in inhomogeneous media 793
Distribution coefficient of indium in germanium on crystallization 60
Domain processes, ferromagnetic, in single crystal disc specimens of silicon–iron (R) 712

Echoes, radio, *see* Radio echoes.
Edge dislocations in inhomogeneous media 793
Editorial 1
Electric strength and plastic deformation of alkali halide crystals (L) 1113
Electric strength of potassium bromide crystals, influence of cathode material (L) 716

PAGE
Electric strength of KCl and NaCl crystals, variation with temperature 74
Electrical breakdown, *see* Breakdown, electrical.
Electrical conductivity and current noise of carbon resistors 680
Electrical conductivity of MgO single crystals at high temperatures 281
Electrical conductivity and thermoelectric power of MgO (L) 612
Electrical discharges, fundamental processes of initiation 665
Electrical properties of molybdenite 377
Electrode contamination in electron optical systems 542
Electrodeless discharge, *see under* Discharge.
Electromagnetic absorption, *see* Absorption, electromagnetic.
Electron currents in thin oxide films on aluminium 533
Electron diffraction, of 150 kv electrons 949
Electron lenses, magnetic, effect of pole piece saturation in 448
Electron microscope lenses, magnetic circuit 441
Electron optical systems, electrode contamination 542
Electrons, penetration into luminescent material 1057
Electrons, secondary emission, origin of 278
Electrons, slow, decomposition of thin films on bombardment with 1073
Electrostatic lenses, correction by departure from rotational symmetry 775
Electrostatic and steady-current problems involving anisotropic bodies 557
Electro-thermal behaviour of point contacts to semiconductors 570
Emission, secondary, *see* Secondary emission.
Energies, β-ray, *see* Beta-ray energies.
Energy balance equation for positive columns of high pressure arcs 169
Equipartition and Johnson noise (L) 714
Excitation, cathode-ray, decay of long period afterglow of alkali halides under 371
Expansion, thermal, measurement of single crystals by photoelectric recording dilatometer 887
Extinction in X-ray diffraction patterns of powders 1003
Extraction and carrier injection in lead sulphide 50

Fabry–Perot interferometer at millimetre wavelengths 597
Fading characteristics, random, some studies on 105
Ferrimagnetic material, thermal effects accompanying magnetization (R) 609
Ferromagnetic domain processes in single crystal disc specimens of silicon–iron (R) 712
Ferromagnetism in pyrrhotite, nature 753
Films, thin aluminium, reflectivity, and use in interferometry 963
Films, thin, decomposition on bombardment with slow electrons 1073
Films, thin oxide, on aluminium, electron currents in 533
Flow, plastic, of iron and plain carbon steels above A_3-point 865
Fluctuations, high frequency, from light sources, study using phototubes and r.f. amplifiers 737
Focusing, vertical, in the microtron (R) 802
Freezing of supercooled water in glass 241
Friction of plastics, effect of temperature 728
Friction of rubber-like materials, influence of load and surface roughness 721
Friction, rubber, surface condition and electric impedance 817
Friction, rubber, velocity and temperature dependence 386

Gamma-line, ^{137}Cs, precision measurement 54
Gas mixtures, viscosity 362
Gases, electrical breakdown, *see under* Breakdown.
Gases, sorption at very low pressures by thorium powder 400
Germanium, impurity concentrations in, method of estimating (R) 984
Germanium, injected absorption 588
Germanium, photoelectromagnetic and photodiffusion effects in 743
Germanium, radiative transitions in (R) 330
Germanium rectifier, *see* Rectifier, germanium.

PAGE
Germanium, single crystal, optical constants (L) 512
Germanium, surface recombination in (L) 899
Grazing-incidence integral Ch (x, χ) for monochromatic absorption in an exponential atmosphere (R) 710
Gutta percha, crystal structures of 7

Hardening, latent, and mean shear stress in array of dislocations 2
Heat flow and growth of metal single crystals from melt 525
Heat, parallel sources, free convection over 393
Heat, specific, *see* Specific heat.
Helium liquefier-cryostat, miniature, of cascade type 296
High-frequency discharges in helium, collisional processes and similarity . . 17
Hydrogen peroxide, formation on fresh metal surfaces 153
Hydrogen spark channels, some properties 500

Illuminating wave, effect of source size on coherence of 975
Images, phase-contrast, theory of (R) 331
Images, *see also* Diffraction images.
Immersion objective, low voltage electron, characteristics 284
Impedance, electrical, and surface condition in rubber friction 817
Impurity concentrations in Ge, method of estimating (R) 984
Impurity, metallic, and closed moulds, effect on growth from melt of single crystals of tin 65
Infra-red absorption spectra, *see under* Absorption.
Infra-red photoconductivity, long wavelength, of silicon at low temperatures . 162
Injecting area contacts on Ge, forward characteristic 841
Injecting point contacts, theory of forward characteristic 833
Injecting properties, carrier, of emitter contacts and light spots at normal and moderately elevated temperatures 845
Injection, carrier, *see* Carrier injection.
Interferometer, Fabry–Perot, at millimetre wavelengths 597
Interferometric studies of growth of stearic acid crystals and their optical properties . . . . 414, corr. 523
Interferometry, use of thin Al films in 963
Ionic current and film growth of thin oxide layers on aluminium 317
Ionized layer with vertical and horizontal ionization gradients, reflection of radio waves from 308
Ionosphere, spread echoes, and recombination and diffusion 1025
Iron oxide powders, interpretation of magnetic properties: II 265
Irradiation, α-particle, and luminescence of air, glass and quartz 81
Irreversible processes, thermodynamics of, application to theory of magnetron (R) . 423

Jet noise, choked, mechanism 1039
Johnson noise and equipartition (L) 714
Joshi effect, negative, and ‘ Corona pressure ’ (L) 256

Latent hardening, *see* Hardening.
Lattices, face-centred cubic, dislocation nodes in 481
Lead, optical constants of PbS, PbSe and PbTe in the $0{\cdot}5$–$3\ \mu$ region of the spectrum 134
Lead sulphide, absorption spectra, at different temperatures 194
Lead sulphide specimens, transistor action and related phenomena in . . . 216
Lead telluride and silicon, inter-relation of optical constants 141
Lenses, objective, for microscopical examination of metals under polarized light, effect of strain in 302
Lenses, *see also* Electron lenses, Electrostatic lenses.
Light sources, high-frequency fluctuations, study using phototubes and r.f. amplifiers 737
Light spots, carrier injecting properties at normal and moderately elevated temperatures 845
Lightning streamers, study with 50 cm radar (R) 895

PAGE
Liquid containing spherical particles, dielectric constant (L) . . . 611
Load, influence on friction of rubber-like materials 721
Low-frequency electrodeless discharge, secondary processes (L) 431
Low pressures, electrical breakdown of gases in non-uniform fields at . . 345
Low temperature conduction in extremely degenerate semiconductors . . . 273
Low temperatures, long wavelength photoconductivity of silicon at . . . 162
Luminescence of air, glass and quartz under α-particle irradiation . . . 81
Luminescent material, penetration of electrons into 1057

Magnesium oxide single crystals at high temperature, electrical conductivity of . 281
Magnesium oxide, thermoelectric power and electrical conductivity . 612
Magnetic circuit in electron microscope lenses 441
Magnetic deflector for synchro-cyclotrons, perturbations in 25
Magnetic electron lenses, effect of pole piece saturation in 448
Magnetic field, spectroscopic study of caesium discharges 33
Magnetic properties of some iron oxide powders, interpretation : II . . . 265
Magnetic viscosity in platinum cobalt (R) . 805
Magnetization of a ferrimagnetic material, thermal effects (R) . . . 609
Magnetization, saturation, of nickel at high pressure (L) 255
Magnetron, application of thermodynamics of irreversible processes to theory of (R) 423
Melt, growth of single crystals of tin from, effect of added impurity and closed moulds on 65
Mercury, thermoelectric effects due to asymmetrical temperature gradients in 649
Mercury, ultra-ionization potentials, and oscillations in space charge detectors 642
Metal layers, condensed from vapour in high vacuum, origin of stress in . 905
Metal plate, effect on total reflection 859
Metal, pure, transient creep in 459
Metal surfaces, fresh, formation of hydrogen peroxide on 153
Meteor trails, decay of radio echoes from (L) 150
Microscope, electron, *see* Electron microscope.
Microscope principle, new 1096
Microtron, phase stability 41
Microtron, vertical focusing in (R) 802
Microtron, 4·5 MeV, design and operation 654
Microwave computing problem (R) 987
Mobility variation, influence in high fields on diffusion theory of rectifier barriers (L) 430
Molybdenite, electrical properties 377

Nickel, saturation magnetization at high pressure (L) 255
Noise, choked jet, mechanism 1039
Noise, rectifier, reduction by illumination (L) 1118
Noise in semiconductors at very low frequencies (L) 259
Noise spectra of point contact Ge rectifier, measurements of 1017
Noise, *see also* Contact noise, Current noise.
Nuclear demagnetization, adiabatic, and nuclear alignment, 1000 kw water-cooled solenoid for experiments on 921

Optical constants of lead sulphide, lead selenide and lead telluride in the 0·5–3 μ region of the spectrum 134
Optical constants for lead telluride and silicon, inter-relation 141
Optical constants of single crystal of germanium (L) 512
Optical properties and growth of stearic acid crystals, further note (L) . . . 989
Optical properties of stearic acid crystals, and interferometric studies of growth . . . . 414, corr. 523
Order–disorder transformation in iron–nickel alloys in the region $FeNi_3$ . . 221
Oscillations in space charge detectors and ultra-ionization potentials of Hg . . 642
Oxide layers, thin, film growth on aluminium, and ionic current 317

PAGE
Particles, spherical, dielectric constant of liquid containing (L) 611
Particles, *see also* Charged particles.
Pentane and alcohol, direct measurement of specific heat at constant volume (R) 421
Phase-contrast images, theory of (R) 331
Phase stability of the microtron 41
Photoconductive and photoelectromagnetic effects in PbS single crystals 993
Photoconductivity, infra-red, *see* Infra-red photoconductivity.
Photodiffusion and photoelectromagnetic effects in Ge 743
Photoelectromagnetic and photoconductive effects in PbS single crystals 993
Photoelectromagnetic and photodiffusion effects in Ge 743
Plastic deformation and electric strength of alkali halide crystals (L) 1113
Plastic flow of iron and plain carbon steels above A$_3$-point 865
Plastics, friction and mechanical properties, effect of temperature 728
Platinum cobalt, magnetic viscosity in (R) 805
Polar liquids, electromagnetic absorption at 3×10^9 c/s 175
Polarization, anodic, structure and orientation of calomel formed on liquid mercury by 129
Polarized light, microscopical examination of metals under, and effect of strain in objective lenses used 302
Pole-piece saturation in magnetic electron lenses 448
Polymers, high, stress, strain and birefringence in 884
Potentials, ultra-ionization of Hg, and oscillations in space charge detectors 642
Powder patterns, extinction in 1003
Propagation of plane sound waves in tubes 769
Propagation of plane sound waves in tubes, theory 695
Proportional-counter technique for measuring x-ray scattering from powders, fibres and liquids 1009
Pyrrhotite, nature of ferromagnetism in 753

Radar, 50 cm, for study of lightning streamers (R) 895
Radiation, radio-frequency, *see* Radio-frequency radiation.
Radio echoes from meteor trails, decay (L) 150
Radio-frequency radiation, solar, thermal theories of high-intensity components 97
Radio waves from ionized layer having both vertical and horizontal ionization gradients 308
Random fading characteristics, some studies on 105
Recombination, surface, in Ge (L) 899
Recombination and diffusion and spread echoes in ionosphere 1025
Rectifier barriers, diffusion theory, influence of mobility variation in high fields (L) 430
Rectifier, cuprous oxide, characteristics 622
Rectifier, germanium, point contact, measurements of noise spectra 1017
Rectifier noise, reduction by illumination (L) 1118
Reflection measurements, determination of infra-red absorption spectra from 969
Reflection, total, effect of metal plate on 859
Reflectivity of thin Al films and their use in interferometry 963
Refractive indices of water vapour, air, oxygen, nitrogen, deuterium and helium 189
Resistors, carbon, electrical conductivity and current noise 680
Rubber friction, surface condition and electrical impedance 817
Rubber friction, velocity and temperature dependence 386
Rubber-like materials, friction of, influence of load and surface roughness 721

Saturation, pole-piece, *see under* Pole-piece saturation.
Scattering, X-ray, from powders, fibres and liquids proportional-counter technique for measuring 1009
Secondary emission electrons, origin of 278
Selenium, amorphous, dielectric constant at wavelengths of 1 cm and 3 cm (L) 1117
Semiconducting compound InSb (L) 1115
Semiconductors, contact noise in (L) 334
Semiconductors, electro-thermal behaviour of point contacts to 570

PAGE
Semiconductors, extremely degenerate, low temperature conduction in . . 273
Semiconductors, noise at very low frequencies 259
Shear rate, variation of transverse viscosity with (L) 614
Shear stress, mean, in array of dislocations and latent hardening 2
Silicon–iron, single crystal disc specimens, ferromagnetic domain processes in (R) . 712
Silicon and lead telluride, inter-relation of optical constants 141
Silicon, long wavelength infra-red photoconductivity at low temperatures . . 162
Similarity and collisional processes in high-frequency discharges in helium . . 17
Solar radio-frequency radiation, *see* Radio-frequency radiation.
Solenoid, 1000 kw water-cooled, for experiments on nuclear alignment and adiabatic nuclear demagnetization, design and construction 921
Solids, amorphous, mechanism of stress birefringence 617
Solids, transparent, stress relaxation in, optical method of observation . . 954
Sorption of gases at very low pressures by thorium powder 400
Sound, propagation of plane waves in tubes 769
Sound, velocity in air at low pressures 760
Sound, velocity of, and viscosity in liquids, relation between 368
Sound waves, plane, theory of propagation in tubes 695
Source size, effect on coherence of illuminating wave 975
Space charge detectors, oscillations in, and ultra-ionization potentials of Hg . . 642
Spark channels, hydrogen, some properties 500
Specific heat at constant volume of pentane and alcohol, direct measurement (R) . 421
Spectra, *see under* Absorption spectra, Noise spectra.
Spectrometer, β-ray, variable-radius semicircular magnetic focusing . . . 911
Spectrometer, short magnetic lens, investigations 937
Spectroscopic study of caesium discharges in magnetic field 33
Spectroscopic study of caesium discharges in magnetic field (L) 898
Stability, phase, *see* Phase stability.
Standing wave method for measuring electromagnetic absorption in polar liquids at frequencies of the order 3×10^9 c/s 175
Steady-current and electrostatic problems involving anisotropic bodies . . 557
Stearic acid crystals, interferometric studies of growth, and their optical properties . . . . 414, corr. 523
Stearic acid crystals, optical properties and growth of, further note (L) . . 989
Steels, iron and plain carbon, plastic flow above A_3-point 865
Stereoscopic vision, role of convergence (L) 513
Strain, effect, in objective lenses used for microscopical examination of metals under polarized light 302
Stress birefringence, mechanism, in amorphous solids 617
Stress, origin, in metal layers condensed from the vapour in high vacuum . . 905
Stress relaxation in transparent solids, optical method of observing . . . 954
Stress, strain and birefringence in high polymers 884
Structure and orientation of calomel formed in liquid mercury by anodic polarization 129
Structure, crystal, *see also* Crystal structure.
Supercooling of water 688
Surface condition and electrical impedance in rubber friction 817
Surface deformation caused on iron crystals by unidirectional abrasion (L) . . 717
Surface, metal, *see* Metal surface.
Surface roughness, influence on friction of rubber-like materials 721
Synchro-cyclotrons, perturbations in magnetic deflector 25

Temperature and variation of electric strength of crystals of KCl and NaCl . . 74
Thermal effects accompanying magnetization of ferrimagnetic material (R) . . 609
Thermal theories of the high intensity components of solar radio-frequency radiation 97
Thermodynamics of irreversible processes, application to theory of magnetron (R) . 423
Thermoelectric effects due to stationary and moving asymmetrical temperature gradients in Hg 649
Thermoelectric power and electrical conductivity of MgO (L) 612

PAGE

Thin films, *see* Films, thin.
Thixotropic fluids, instantaneous, comparison of absolute viscosity measurements on 120
Thixotropic materials, measurement of time-dependence of viscosity of 115
Thorium powder, sorption of gases at very low pressures 400
Transient creep in pure metals 459
Transistor action and related phenomena in lead sulphide specimens from various sources 216
Transistors, silicon point-contact, properties (L) 429
Transitions, radiative, in germanium (R) 330

Vapour pressure ratio of $^{12}C^{16}O$ and $^{13}C^{16}O$ (L) 808
Velocity, sound, *see* Sound, velocity of.
Viscosity, absolute, of anomalous fluids, measurement:
  I—Time dependence 115
  II—Comparison of absolute viscosities 120
Viscosity of gas mixtures 362
Viscosity in liquids and velocity of sound, relation between 368
Viscosity, magnetic, in platinum cobalt (R) 805
Viscosity, transverse, variation with shear rate (L) 614
Vision, stereoscopic, role of convergence (L) 513

Water bells 1067
Water, supercooled, freezing in glass 241
Water, supercooling 688
Waves, radio, *see* Radio waves.
Wear process, quantitative study 929

X-ray diffraction patterns of powders, extinction in 1003
X-ray scattering, *see under* Scattering.

INDEX OF AUTHORS (WITH TITLES)

PAGE

Agarwela, R. P., and Wilman, H.: The surface deformation caused on iron crystals by unidirectional abrasion (L) . . . 717
Alexopoulos, K., and Theodossiou, A.: On the nature of ferromagnetism in pyrrhotite . . . 753
Allen, N. L.: An experiment on the radial motion of charged particles from a pulsed arc discharge . . . 1087
Arndt, U. W., Coates, W. A., and Riley, D. P.: A proportional-counter technique for measuring x-ray scattering from powders, fibres and liquids . . . 1009
Arnikar, H. J.: The 'corona pressure' and negative Joshi effect (L) . . . 256
Avery, D. G.: The optical constants of lead sulphide, lead selenide and lead telluride in the $0{\cdot}5$–$3\,\mu$ region of the spectrum . . . 134
Avery, D. G., and Clegg, P. L.: The optical constants of a single crystal of germanium . . . 512

Baines, W. D., *see* Rouse, Hunter.
Baker, L. R.: The effect of source size on the coherence of an illuminating wave . . . 975
Bakhshi, N. N., *see* Parthasarathy, S.
Banbury, P. C.: Carrier injection and extraction in lead sulphide . . . 50
Banbury, P. C.: Theory of the forward characteristic of injecting point contacts . . . 833
Banerjee, M. K., and Saha, A. K.: Investigations on a short magnetic lens spectrometer . . . 937
Banerji, R. B.: Some studies on random fading characteristics . . . 105
Bardsley, W., *see* Granville, J. W.
Bates, L. F., and Sherry, N. P. R.: Thermal effects accompanying magnetization of a ferrimagnetic material (R) . . . 609
Bell, D. A.: Johnson noise and equipartition (L) . . . 714
Bell, J. S.: Vertical focusing in the microtron (R) . . . 802
Bhatawdekar, M. G., *see* Dogra, O. P.
Bigg, E. K.: The supercooling of water . . . 688
Birks, J. B., and King, J. W.: The luminescence of air, glass and quartz under α-particle irradiation . . . 81
Bose, H. N., and Sharma, J.: Decay of long period afterglow of alkali halides under cathode ray excitation . . . 371
Bracewell, R. N.: A microwave computing problem (R) . . . 987
Braybon, J. E. H.: The mechanism of stress birefringence in amorphous solids . . . 617
Bremner, J. W.: The variation of transverse viscosity with shear rate (L) . . . 614
Bryant, M. O., and Jones, G. O.: Direct measurement of the specific heat at constant volume of pentane and alcohol . . . 421
Burfoot, J. C.: Correction of electrostatic lenses by departure from rotational symmetry . . . 775
Burgess, R. E.: Contact noise in semiconductors (L) . . . 334
Burgess, R. E.: The influence of mobility variation in high fields on the diffusion theory of rectifier barriers . . . 430
Burridge, J. C., Kuhn, H., and Pery, Anne: Reflectivity of thin aluminium films and their use in interferometry . . . 963

Cade, R.: Some electrostatic and steady-current problems involving anisotropic bodies . . . 557
Calderwood, J. H., and Cooper, R.: Variation of the electric strength of KCl and NaCl crystals with temperature . . . 74
Campbell, C. G., and Kyles, J.: The variable-radius semicircular magnetic focusing β-ray spectrometer . . . 911
Campbell, I. D., *see* Weston, D. E.
Caro, D. E., and Martin, L. H.: The velocity of sound in air at low pressures . . . 760
Champion, K. S. W.: The energy balance equation for the positive columns of high pressure arcs . . . 169

PAGE

Chapman, S. : Note on the grazing-incidence integral Ch (x, χ) for monochromatic absorption in an exponential atmosphere (R) 710
Charlesby, A. : Electron currents in thin oxide films on aluminium 533
Charlesby, A. : Ionic current and film growth of thin oxide layers on aluminium . 317
Chester, P. F., and Jones, G. O. : A miniature helium liquefier-cryostat of cascade type 296
Clarke, W. W. H., and Jacob, L. : Some characteristics of a low voltage electron immersion objective 284
Clegg, P. L., *see* Avery, D. G.
Coates, W. A., *see* Arndt, U. W.
Cooper, R., *see* Calderwood, J. H.
Cooper, R., and Grossart, D. T. : The influence of cathode material on the electric strength of potassium bromide crystals (L) 716
Cooper, R., and Wallace, A. A. : Plastic deformation and the electric strength of alkali halide crystals (L) 1113
Cowley, J. M. : A new microscope principle 1096
Craggs, J. D., *see* Craig, R. D.
Craig, H., and Dietrich, C. F. : Development of a new method for the absolute determination of β-ray energies. 201
Craig, R. D., and Craggs, J. D. : Some properties of hydrogen spark channels. . 500
Crawford, S. M. : An optical method of observing stress relaxation in transparent solids. 954
Crawford, S. M. : The relation between stress, strain and birefringence in some high polymers 884
Culshaw, W. : The Fabry–Perot interferometer at millimetre wavelengths . . 597
Culshaw, W., and Jones, D. S. : Effect of a metal plate on total reflection . . 859
Cunnell, F. A., Saker, E. W., and Edmond, J. T. : A note on the semiconducting compound InSb (L) 1115

Daniels, J. M. : The design and construction of a 1000 kw water cooled solenoid intended for experiments on nuclear alignment and adiabatic nuclear demagnetization 921
Davies, L. W. : Spectroscopic study of caesium discharges in a magnetic field. . 33
Davies, L. W. : Spectroscopic study of caesium discharges in a magnetic field (L) . 898
Denny, D. F. : The influence of load and surface roughness on the friction of rubber-like materials 721
Dietrich, C. F., *see* Craig, H.
Dogra, O. P., Bhatawdekar, M. G., and Ramaiah, N. A. : A note on secondary processes in a low-frequency electrodeless discharge 431
Dowd, J. J., and Rouse, R. L. : Distribution coefficient of indium in germanium on crystallization (L) 60

Eckersley, T. L. : Recombination and diffusion and spread echoes in the ionosphere 1025
Edmond, J. T., *see* Cunnell, F. A.
Ehrenberg, W., and Franks, J. : The penetration of electrons into luminescent material. 1057
El-Sabeh, S. H. M., and Hasted, J. B. : The dielectric constant of a liquid containing spherical particles (L) 611
Essen, L. : The refractive indices of water vapour, air, oxygen, nitrogen, hydrogen, deuterium and helium 189

Feltham, P. : The plastic flow of iron and plain carbon steels above the A_3-point . 865
Finch, G. I., Lewis, H. C., and Webb, D. P. D. : The diffraction of 150 kv electrons 949
Fisher, D. : Crystal structures of gutta percha 7
Franks, J., *see* Ehrenberg, W.
Fuschillo, N. : Thermoelectric effects due to stationary and moving asymmetrical temperature gradients in mercury 649

PAGE
Gibson, A. F.: Injected absorption in germanium 588
Gibson, A. F.: *see also* Granville, J. W.
Goodwin, D. W., and Macfadyen, K. A.: Electrical conduction and breakdown in liquid dielectrics . . . . 85, corr. 815
Goss, A. J.: Heat flow and the growth of metal single crystals from the melt . . 525
Goss, A. J.: The effect of added metallic impurity and of closed moulds on the growth from the melt of single crystals of tin 65
Granville, J. W., and Bardsley, W.: Some properties of silicon point-contact transistors 429
Granville, J. W., and Gibson, A. F.: The reduction of rectifier noise by illumination (L) 1118
Greenhow, J. S., *see* Kaiser, T. R.
Grossart, D. T., *see* Cooper, R.
Grunberg, L.: The formation of hydrogen peroxide on fresh metal surfaces . . 153
Gunn, J. B.: Radiative transitions in germanium (R) 330

Haines, H. R., *see* Mott, B. W.
Harcombe, D., *see* Morgan, C. Grey
Hasted, J. B., *see* El-Sabeh, S. H. M.
Head, A. K.: Edge dislocations in inhomogeneous media 793
Heath, H. R.: The viscosity of gas mixtures 362
Henderson, C., Heymann, F. F., and Jennings, R. E.: Phase stability of the microtron 41
Henderson, C., Heymann, F. F., and Jennings, R. E.: The design and operation of a 4·5 Mev microtron 654
Henderson, F., *see* Wain, H. L.
Henisch, H. K., *see* Tipple, P. M.
Henisch, H. K., and Morten, F. D.: Forward characteristic of injecting area contacts on germanium 841
Hewitt, F. J.: The study of lightning streamers with 50 cm radar (R) . . . 895
Heymann, F. F., *see* Henderson, C.
Hogarth, C. A.: A study of carrier injecting properties of emitter contacts and light spots at normal and moderately elevated temperatures 845
Hogarth, C. A.: Transitor action and related phenomena in lead sulphide specimens from various sources 216
Hopkins, H. H.: A note on the theory of phase-contrast images (R) . . . 331
Humphreys, H. W., *see* Rouse, Hunter
Hyde, F. J.: Measurements of noise spectra of a point contact germanium rectifier . 1017

Jacob, L., *see* Clarke, W. W. H.
Jennings, R. E., *see* Henderson, C.
Johns, T. F.: Vapour pressure ratio of $^{12}C\,^{18}O$ and $^{13}C\,^{16}O$ (L) 808
Jones, D. S., *see* Culshaw, W.
Jones, F. Llewellyn, and Williams, G. C.: Collisional processes and similarity in high-frequency discharges in helium. 17
Jones, F. Llewellyn, and Williams, G. C.: The electrical breakdown of gases in non-uniform fields at low pressure 345
Jones, G. O., *see* Bryant, M. O., Chester, P. F.
Jones, G. O., and Stacey, F. D.: The saturation magnetization of nickel at high pressure (L) 255
Jones, R. V., *see* Paul, W.

Kaiser, T. R., and Greenhow, J. S.: On the decay of radio echoes from meteor trails (L) 150
King, J. W., *see* Birks, J. B.
King, R. F., and Tabor, D.: The effect of temperature on the mechanical properties and the friction of plastics. 728

PAGE
Klinger, Y., and Saker, E. W. : The dielectric constant of amorphous selenium at wavelengths of 1 cm and 3 cm (L) 1117
Kuhn, H., *see* Burridge, J. C.
Kyles, J., *see* Campbell, C. G.

Lance, G. N., and Perry, R. L. : Water bells 1067
Landecker, K., and Robinson, B. J. : Study of high-frequency fluctuations from light sources using phototubes and tuned radio-frequency amplifiers 737
Lang, A. R. : Extinction in x-ray diffraction patterns of powders 1003
Le Couteur, K. J. : Perturbations in the magnetic deflector for synchro-cyclotrons . 25
Lees, J. : Cuprous oxide rectifier characteristics. 622
Lempicki, A. : The electrical conductivity of MgO single crystals at high temperatures 281
Lempicki, A. : The origin of secondary emission electrons 278
Lewis, H. C., *see* Finch, G. I.
Lewis, T. J. : The dependence of the dielectric strength of pure liquids on cathode material (R) 425
Liebmann, G. : The effect of pole piece saturation in magnetic electron lenses . 448
Lindsay, P. A., and Sims, G. D. : Application of the thermodynamics of irreversible processes to the theory of the magnetron 423
Lindstrom, G., Siegbahn, K., and Wapstra, A. H. : A precision measurement of the ^{137}Cs γ-line 54
Linfoot, E. H., and Wolf, E. : Diffraction images in systems with an annular aperture 145
Little, V. I. : An analogue computer employing the principle of the Kelvin bridge 185
Little, V. I. : A standard wave method for measuring electromagnetic absorption in polar liquids at frequencies of the order 3×10^9 c/s 175

MacDonald, D. K. C., *see* Templeton, I. M.
MacFadyen, K. A., *see* Goodwin, D. W.
Mansfield, R. : The electrical conductivity and thermoelectric power of magnesium oxide 612
Mansfield, R., and Salam, S.A. : Electrical properties of molybdenite . . 377
Martin, D. H. : Ferromagnetic domain processes in single crystal disc specimens of silicon iron (R) 712
Martin, L. H., *see* Caro, D. E.
Morgan, C. Grey, and Harcombe, D. : Fundamental processes of the initiation of electrical discharges 665
Morten, F. D., *see* Henisch, H. K.
Moss, T. S. : Inter-relation between optical constants for lead telluride and silicon 141
Moss, T. S. : Photoelectromagnetic and photoconductive effects in lead sulphide single crystals. 993
Moss, T. S., Pincherle, L., and Woodward, A. M. : Photoelectromagnetic and photodiffusion effects in germanium 743
Mott, B. W., and Haines, H. R. : The effect of strain in objective lenses used for microscopical examination of metals under polarized light 302
Mulvey, T. : The magnetic circuit in electron microscope lenses 441
Murbach, H. P., and Wilman, H. : The origin of stress in metal layers condensed from the vapour in high vacuum 905
Myers, H. P. : A simple varying capacitor method for the measurement of contact potential difference in high vacuum 493

Newhouse, V. L., *see* Tebble, R. S.

Ogle, K. N.: Wright, W. D. : The role of convergence in stereoscopic vision . 513
Osmond, W. P. : An interpretation of the magnetic properties of some iron oxide powders : II 265

PAGE

Parthasarathy, S., and Bakhshi, N. N. : Relation between velocity of sound and viscosity in liquids 368
Paul, W., and Jones, R. V. : Absorption spectra of lead sulphide at different temperatures 194
Perry, R. L., *see* Lance, G. N.
Pery, Anne, *see* Burridge, J. C.
Piddington, J. H. : Thermal theories of the high-intensity components of solar radio-frequency radiation. 97
Pincherle, L., *see* Moss, T. S.
Poole, K. M. : Electrode contamination in electron optical systems 542
Powell, A. : On the mechanism of choked jet noise 1039
Price, W. C., *see* Robinson, T. S.

Rabinowicz, E. : A quantitative study of the wear process 929
Rae, D., *see* Thornton, S.
Ramaiah, N. A., *see* Dogra, O. P.
Reynolds, P. M., *see* Verma, Ajit Ram.
Reynolds, W. N. : Surface recombination in germanium (L) 899
Riley, D. P., *see* Arndt, U. W.
Robinson, B. J., *see* Landecker, K.
Robinson, T. S., and Price, W. C. : The determination of infra-red absorption spectra from reflection measurements 969
Rollin, B. V., and Simmons, E. L. : Long wavelength infra-red photoconductivity of silicon at low temperatures 162
Rollin, B. V., and Templeton, I. M. : Noise in semiconductors at very low frequencies (L) 259
Rose, F. W. G., and Timmins, E. W. : A method of estimating impurity concentrations in germanium (R) 984
Rouse, Hunter, Baines, W. D., and Humphreys, H. W. : Free convection over parallel sources of heat 393
Rouse, R. L., *see* Dowd, J. J.

Saha, A. K., *see* Banerjee, M. K.
Saker, E. W., *see* Cunnell, F. A.
Saker, E. W., *see* Klinger, Y.
Salam, S. A., *see* Mansfield, R.
Schallamach, A. : Surface condition and electrical impedance in rubber friction . 817
Schallamach, A. : The velocity and temperature dependence of rubber friction . 386
Sharma, J., *see* Bose, H. N.
Sherry, N. P. R., *see* Bates, L. F.
Siegbahn, K., *see* Lindstrom, G.
Simmons, E. L., *see* Rollin, B. V.
Simpson, A. W., and Tredgold, R. H. : Magnetic viscosity in platinum cobalt (R) . 805
Sims, G. D., *see* Lindsay, P. A.
Spells, K. E. : Sensitive bubble jets (L) 258
Stacey, F. D., *see* Jones, G. O.
Stroh, A. N. : The mean shear stress in an array of dislocations and latent hardening 2
Stuckes, A. D. : Electro-thermal behaviour of point contacts to semiconductors . 570

Tabor, D., *see* King, R. F.
Tebble, R. S., and Newhouse, V. L. : The Barkhausen effect in single crystals . 633
Templeton, I. M., *see* Rollin, B. V.
Templeton, I. M., and MacDonald, D. K. C. : The electrical conductivity and current noise of carbon resistors 680
Theodossiou, A., *see* Alexopoulos, K.
Thirsk, H. R. : The structure and orientation of calomel formed on liquid mercury by anodic polarization 129
Thompson, N. : Dislocation nodes in face-centred cubic lattices 481

PAGE

Thornton, S. : The measurement of the absolute viscosity of anomalous fluids : I—the measurement of the time-dependence of viscosity of thixotropic materials . 115
Thornton, S., and Rae, D. : The measurement of the absolute viscosity of anomalous fluids : II—a comparison of absolute viscosity measurements on instantaneously thixotropic fluids 120
Timmins, E. W., *see* Rose, F. W. G.
Tipple, P. M., and Henisch, H. K. : Thermal effects at point contact diodes . . 826
Tredgold, R. H., *see* Simpson, A. W.
Trezona, P. W. : Additivity of colour equations 548

Verma, Ajit Ram, and Reynolds, P. M. : A further note upon the growth and optical properties of stearic acid crystals (L) 989
Verma, Ajit Ram, and Reynolds, P. M. : Interferometric studies of the growth of stearic acid crystals and their optical properties . . . . . 414, corr. 523
Vernon, E. V., and Weintroub, S. : The measurement of the thermal expansion of single crystals of indium and tin with a photoelectric recording dilatometer . 887

Wagener, S. : Sorption of gases at very low pressures by thorium powder . . 400
Wain, H. L., and Henderson, F. : Room temperature brittleness of chromium. . 515
Wakelin, R. J., and Yates, E. L. : A study of the order–disorder transformation in iron–nickel alloys in the region $FeNi_3$ 221
Waldo Lewis, R. P. : The reflection of radio waves from an ionized layer having both vertical and horizontal ionization gradients 308
Wallace, A. A., *see* Cooper, R.
Wapstra, A. H., *see* Lindstrom, G.
Webb, D. P. D., *see* Finch, G. I.
Weintroub, S., *see* Vernon, E. V.
Weston, D. E. : The theory of the propagation of plane sound waves in tubes. . 695
Weston, D. E., and Campbell, I. D. : Experiments on the propagation of plane sound waves in tubes : I—the adiabatic region. II—the transition region . 769
White, O. M. : Oscillations in space charge detectors and the ultra-ionization potentials of mercury 642
Williams, G. C., *see* Jones, F. Llewellyn.
Wilman, H., *see* Agarwala, R. P., Minbach, H. P.
Wolf, E., *see* Linfoot, E. H.
Woods, J., *see* Wright, D. A.
Woodward, A. M., *see* Moss, T. S.
Wright, D. A., and Woods, J. : The decomposition of thin films on bombardment with slow electrons 1073
Wright, R. W. : Low temperature conduction in extremely degenerate semiconductors 273
Wright, W. D., *see* Ogle, K. N.: Wright, W. D.
Wyatt, O. H. : Transient creep in pure metals 459
Wylie, R. G. : The freezing of supercooled water in glass 241

Yates, E. L., *see* Wakelin, R. J.

INDEX TO REVIEWS OF BOOKS

PAGE
Batchelor, G. K. : *The Theory of Homogeneous Turbulence* 812
Bertele, H. : *Niederdruck-Stromrichterventile* 521
Bull, Henry B. : *Physical Biochemistry* 1126

Calthrop, J. E. : *Advanced Experiments in Practical Physics* 438
Cottrell, A. H. : *Dislocations and Plastic Flow in Crystals* 991
Courty, C. : *Charbons activés* 342

Den Hartog, J. P. : *Advanced Strength of Materials* 436
Dow, W. G. : *Fundamentals of Engineering Electronics* 517

Françon, M. : *Le constraste de phase en optique et en microscopie* 62

Glaser, Walter : *Grundlagen der Elektronenoptik* 810
Green, H. S. : *Molecular Theory of Fluids* 437

Harkins, W. D. : *The Physical Chemistry of Surface Films* 1126
Hartree, D. R. : *Numerical Analysis* 435
Heywood, R. B. : *Designing by Photoelasticity* 523
Holton, G. : *Introduction to Concepts and Theories in Physical Science* 335
Hume-Rothery, W., Christian, J. W., and Pearson, W. B. : *Metallurgical Equilibrium Diagrams* 341
Hynek, J. A. (Ed.) : *Astrophysics. A Topical Symposium* 343

Ingerslev, F. : *Acoustics in Modern Building Practice* 62

Knoll, M., and Kazan, B. : *Storage Tubes and Their Basic Principles* 615
Korn, G. A., and Korn, T. M. : *Electronic Analog Computers* 340

Lloyd, Ll. S. : *Music and Sound* 63

Martin, L. C. : *Technical Optics, Volume II* 62
Massey, H. S. W., and Burhop, E. H. S. : *Electronic and Ionic Impact Phenomena* 434
Mataré, H. F. : *Empfangsprobleme im Ultrahochfrequenzgebiet* 61
Milne, E. A., the late : *Sir James Jeans—A Biography* 336
Mole, J. H. : *Filter Design Data for Communication Engineers* 438
Murray, H. D. (Ed.) : *Colour in Theory and Practice* 261
Murray, W. M. (Ed.) : *Fatigue and Fracture of Metals* 439

Neblette, C. B. : *Photography—Its Materials and Processes* 990
Nickson, J. J. (Ed.) : *Symposium on Radiobiology : the Basic Aspects of Radiation Effects on Living Systems* 338

Prandtl, L. : *The Essentials of Fluid Dynamics* 518

Ramsauer, Carl : *Grundversuche der Physik in historischer Darstellung, Vol. I, Von den Fallgesetzen bis zu den electrischen Wellen* 814
Reimann, Arnold L. : *Vacuum Technique.* 902
Rosenhead, L., and others : *A Selection of Tables for use in Calculations of Compressible Airflow* 520
Rosser, J. Barkley : *Logic for Mathematicians* 1124

Sauer, R. : *Anfangswertprobleme bei partieller Differentialgleichungen* 437
Scott, J. F. : *The Scientific Work of René Descartes* (1596–1650) 433
Shoenberg, D. : *Superconductivity* 901
Sutton, O. G. : *Micrometeorology* 903

Wilson, E. Bright : *An Introduction to Scientific Research* 522
Winstanley, J. W. : *Text Book on Sound* 62